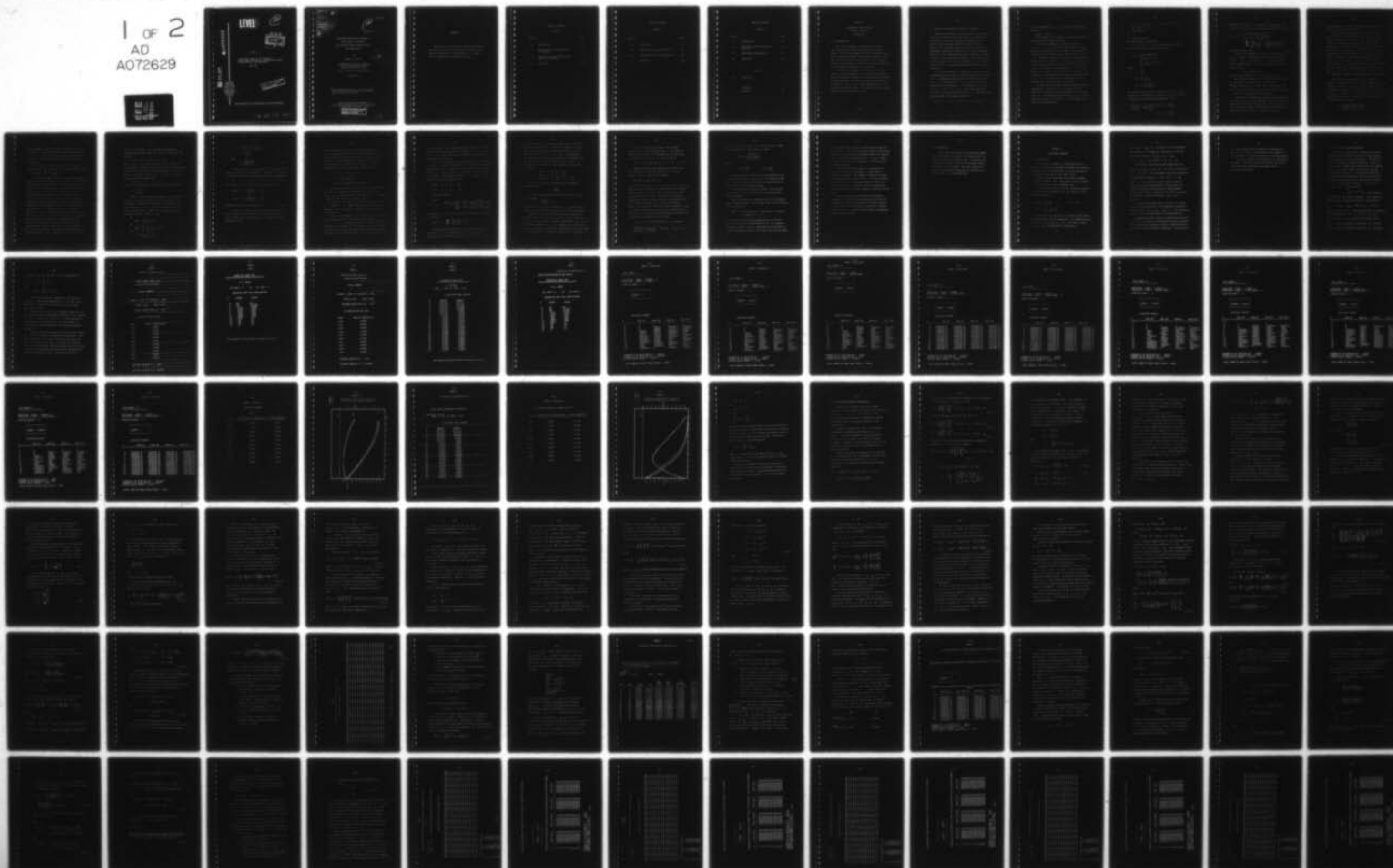
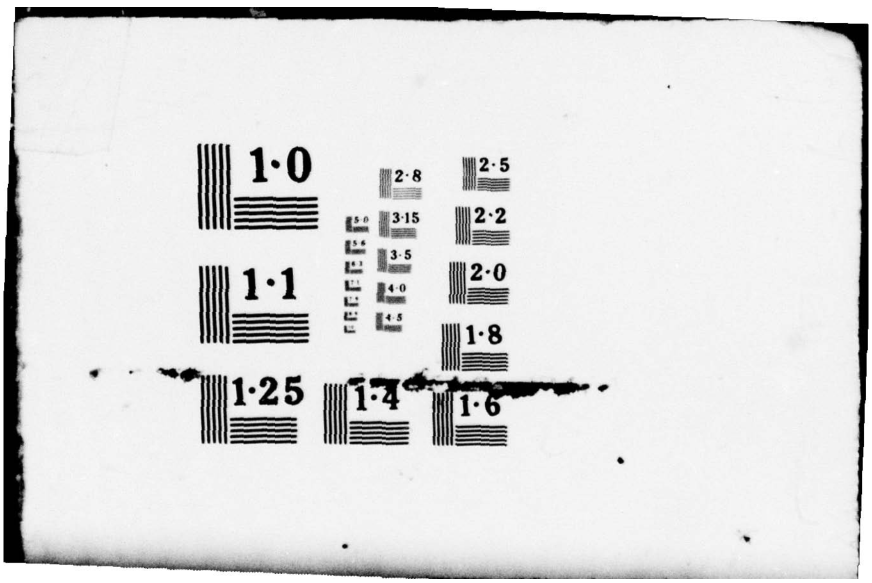
AD-A072 629 UNION COLL AND UNIV SCHENECTADY NY INST OF ADMINISTR--ETC F/G 12/1
SEQUENTIAL ANALYSIS OF VARIANCE: MONTE CARLO SIMULATION, MULTIV--ETC(U)
AUG 79 R W MILLER N00014-77-C-0438
UNCLASSIFIED AES-7907 NL

1 OF 2
AD
A072629




1·0
2·8
2·5
5·0
3·15
2·2
5·6
3·5
2·0
1·1
4·0
4·5
1·8
1·25
1·4
1·6

See 1473 in back

SEQUENTIAL ANALYSIS OF VARIANCE:
MONTE CARLO SIMULATION, MULTIVARIATE NORMAL
APPROXIMATION, AND ROBUSTNESS

AES-7907

ADMINISTRATIVE AND ENGINEERING SYSTEMS MONOGRAPH

AES-7907

(12)

SEQUENTIAL ANALYSIS OF VARIANCE: MONTE CARLO SIMULATION, MULTIVARIATE NORMAL APPROXIMATION, AND ROBUSTNESS*

by

Robert W. Miller

Prepared with partial support
of the Office of Naval Research
Contract N00014-77-C-0438

Leo A. Aroian, Project Director

August 1979

*This monograph contains the first, third, fourth and fifth chapters of the doctoral thesis of Robert W. Miller

UNION COLLEGE AND UNIVERSITY
Institute of Administration and Management
Schenectady, New York 12308

ABSTRACT

Monte Carlo simulation and a multivariate normal approximation are used to find the OC and ASN of the sequential analysis of variance test. It is shown that the SANOVA test is remarkably robust.

TABLE OF CONTENTS

CHAPTER 1

TABLE OF CONTENTS

CHAPTER 3

TABLE OF CONTENTS

CHAPTER 4

CHAPTER 5

CHAPTER 1

SEQUENTIAL FIXED EFFECTS ONE-WAY ANALYSIS OF VARIANCE

1.0 INTRODUCTION

This first chapter of the thesis will consider both the fixed and sequential analysis of variance tests. For the fixed sample test the discussions consist of the statistical model, the optimum properties of the test, and the operating characteristic (OC) function. Each of these concepts is important for the consideration of the sequential analysis of variance test.

The sequential analysis of variance test (termed SANOVA) is first discussed from a historical perspective. Further discussions consist of the experimental procedure, the test statistic, and the test statistic decision rule or regions. The OC and average sample number (ASN) functions are also defined. These functions are extremely helpful for designing SANOVA tests.

1.1 ONE-WAY FIXED EFFECTS ANALYSIS OF VARIANCE

Analysis of variance, a term introduced into statistics by R.A. Fisher (1918, 1925, 1935), is a statistical technique for analyzing measurements depending upon several kinds of effects operating simultaneously. In general, this technique consists of a body of tests of hypotheses, methods of estimation, etc., using statistics which are linear combinations of sums of squares of linear functions of the observed measurements. The simplest case in which analysis of variance is applied, is the one-way classification, in which the observations depend upon only one factor.

In the one-way layout, a population is stratified into m subpopulations according to some characteristic or factor and n_i independent observations are taken from each of k of the m subpopulations $(i = 1,...,k)$. Let the jth observation from population i be denoted by x_{ij} where $i = 1,...,k$ and $j = 1,...,n_i$. Given that population i has mean $\mu+\sigma_i$ and standard deviation σ_i, the statistical model employed in the one-way layout is

$$x_{ij} = \mu + \sigma + e_{ij} \ , \quad i = 1,\cdots,k; \quad j = 1,\cdots,n_i$$

with the parameters $\delta_1,\ldots,\delta_k$ satisfying the following condition

$$n_1\delta_1+\ldots+n_k\delta_k = 0$$

The parameter δ_i is referred to as the differential effect due to the factor at level i.

The usual hypothesis of interest is whether $\delta_1 = \delta_2 = .. = \delta_k = 0$, which is equivalent to the hypothesis of the equality of the k means. The analysis of the effect of the factor depends upon whether $k < m$ or $k = m$. Eisenhardt (1947) was the first to differentiate between the two situations. He used the terms Model I or a fixed effects model as the case where the sample consists of all groups in the population, i.e., $k = m$, and Model II or a random effects model as the case where the interest is in the population from which the sample came, i.e., $k < m$. This thesis will be concerned with only fixed-effects one-way analysis of variance.

The analysis of variance technique requires several assumptions. Specifically, it is assumed that the observations from each of the subpopulations are random variables distributed normally with mean $\mu + \delta_i$ and standard deviation $\sigma = \sigma_i$ for all i. In other words, the model may be expressed as

$$x_{ij} = \mu + \delta_i + e_{ij} \qquad i = 1,\ldots,k;\ j = 1,\ldots,n$$

$$x_{ij} \sim N(\mu + \delta_i,\ \sigma)$$

$$e_{ij} \sim N(0,\ \sigma)$$

and

$$\mathrm{cov}(x_{ij},\ x_{\ell m}) = 0.$$

With this model the hypotheses

H_0: $\mu_1 = \mu_2 = \cdots \mu_k$ vs. H_1: not all means equal

can be tested with the following statistic

$$F_{cal} = \frac{\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{\bar{x}})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - x_i)^2/(N-k)}$$

where

$$N = \sum_{i=1}^{k} n_i$$

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

$$\bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$$

This statistic can be shown (Kempthorne, 1952) to be distributed as a noncentral F variate with (k-1, N-k) degrees of freedom and noncentrality parameter $\bar{n}\lambda$, where

$$\lambda = \frac{\sum_{i=1}^{k}\delta_i{}^2 n_i}{\sigma^2} = \frac{\sum_{i=1}^{k}(\mu_i - \bar{\mu})^2 n_i}{\sigma^2} \quad \text{with} \quad \bar{\mu} = \frac{1}{k}\sum_{i=1}^{k}\mu_i$$

$$\text{and} \quad \bar{n} = \frac{1}{k}\sum_{j=1}^{k} n_i$$

The density function of a noncentral F variate with v_1, v_2 degrees of freedom and noncentrality parameter λ is given by:

$$f_{v_1,v_2,\lambda}(x) = \frac{e^{-\frac{1}{2}\lambda}\; v_1^{\frac{1}{2}v_1}\; v_2^{\frac{1}{2}v_2}\; x^{\frac{1}{2}v_1-1}}{B(\frac{1}{2}v_1, \frac{1}{2}v_2)\;(v_2+v_1x)^{\frac{1}{2}(v_1+v_2)}}$$

$$\sum_{j=0}^{\infty} \left[\frac{\frac{1}{2}\lambda v_1 x}{v_2 + v_1 x}\right]^j \frac{\Gamma\{\frac{1}{2}(2j + v_1 + v_2)\}}{j!\,\Gamma(\frac{1}{2}v_2)\;\Gamma\frac{1}{2}(2j + v_1)}$$

(Johnson and Kotz, 1970). (1.1.1)

If the null hypothesis is true, the distribution of F_{cal} is a central F distribution with k-1, N-k degrees of freedom. Hence, if the hypothesis is rejected whenever F_{cal} is greater than the $100(1-\alpha)\%$ point of this distribution, that is

$$F_{cal} > F^*_{k-1,N-k,1-\alpha}$$

then the significance level of the test will be α.

The operating characteristic curve of the test, that is, the probability of accepting H_0 is given by $\Pr\{F_{cal} \le F^*_{k-1,N-k,1-\alpha}\}$. Since $F_{cal} \sim F_{k-1,N-k,\bar{n}_\xi}$ the OC of the test is characterized by the parameter $\xi = \bar{n}\lambda$, i.e.

$$OC(\lambda) = \Pr\{F_{k-1,N-k,\xi} \le F^*_{k-1,N-k,1-\alpha}\}$$

Several sets of tables and curves have been prepared from which the OC curve for selected tests can be obtained (Tang 1938, Pearson and Hartley 1951, Lehmer 1944, Fox 1956, Fix 1949). Most of these tables are entered with a different parameter than ξ. Appendix A contains a

computer program which will calculate the OC curve (as a function of λ) for any given test.

Originally ANOVA was derived from a distributional point of view, but the F-test has been found to possess several optimum properties. Hsu (1941) showed that the F-test is UMP amongst all tests of size α whose power depends upon λ, and Wald (1942a) proved that the F-test is best when one is interested uniformly in all alternatives, as expressed by uniform weighting on spheres. As far as ANOVA is concerned it is immaterial whether the value of λ is built up by a number of small contributions or a single large one. Situations where instead the main emphasis is on detection of large deviations should not use ANOVA since the test is no longer optimum in these cases.

1.2 SEQUENTIAL ONE-WAY FIXED EFFECTS ANALYSIS OF VARIANCE

Wald (1947) first presented, and systematically studied, the sequential test of a simple hypothesis against a simple alternative. Let H_0 denote the hypothesis that the population density is $f_0(x)$, and H_1 the hypothesis that it is $f_1(x)$. Constants A and B are chosen ($A > B$), and after each observation in a sequence the corresponding likelihood ratio is computed:

$$\Lambda_n = \frac{f_1(x_1)\cdot f_1(x_2)\cdots f_1(x_n)}{f_0(x_1)\cdot f_0(x_2)\cdots f_0(x_n)}$$

The procedure is then as follows: reject H_0 if $\Lambda_n \geq A$, accept H_0 if $\Lambda_n \leq B$, and obtain another observation if $B < \Lambda_n < A$. A and B are chosen so as to make the probabilities of Type-I and Type-II errors equal to α and β respectively.

Exact values of A and B are difficult to obtain. However, Wald (1947) proved that for small values of α and β

$$A \simeq \frac{1 - \beta}{\alpha} \quad \text{and} \quad B \simeq \frac{\beta}{1 - \alpha}$$

Since the hypothesis about the equality of K normal population means with common unknown variance is a composite multiparameter hypothesis with a nuisance parameter, Wald's theory of the sequential probability ratio test cannot be directly applied. To deal with problems such as these, Wald introduced the method of weight functions which, through the notion of a prior distribution for unknown parameters, essentially reduced the basic problem to test hypotheses in one parameter families. A difficulty with this procedure is the choice of the weight function. Cox (1952) devised a unified method under which sequential tests can be obtained for composite hypotheses. The basic idea behind Cox's procedure is to consider a sequence formed by transforming the original observations, the transformation chosen so that the new sequence depends upon a single parameter. Although the distribution of the transformed values $\{T_n\}$ depends upon only a single para-

meter Θ, the sequence $\{T_n\}$ may not be independent. Cox gave conditions under which the following factorization is possible

$$f(T_1, T_2, \cdots, T_n) = f(T_n|\Theta) f(T_2, \cdots, T_n)$$

where $f(T_2, \cdots, T_n)$ does not depend upon Θ. When this factorization is possible a sequential test can be developed to make a decision about this single parameter Θ, using only the transformed values $\{T_n\}$. The test for discriminating between the hypotheses

$$H_0: \ \Theta = \Theta_o \quad \text{vs.} \quad H_1: \Theta = \Theta_1$$

can now be constructed by considering the following ratio

$$\Lambda_n = \frac{f(T_n|\Theta_1)}{f(T_n|\Theta_0)} .$$

Johnson (1953) applied Cox's method to the following one-way fixed effects analysis of variance problem. An experiment is carried out in stages, and at each stage a fixed number r_i, for $i = 1, \cdots, k$, of observations are taken from each group. Denote the jth observation on the ith group at the nth stage by X_{ijn}.

Let

$$SSB_n = n \sum_{i=1}^{k} r_i (\bar{X}_i - \bar{\bar{X}})^2$$

and

$$SSW_n = \sum_{i=1}^{k} \sum_{j=1}^{r_i} \sum_{s=1}^{n} (X_{ijs} - \bar{X}_i)^2$$

with

$$\bar{X}_i = \frac{1}{nr_i} \sum_{j=1}^{n_i} \sum_{s=1}^{r} X_{ijs}$$

$$\bar{\bar{X}} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \sum_{s=1}^{r} X_{ijs}$$

$$N = n \sum_{i=1}^{k} r_i$$

and

$$F_n = \frac{SSB_n/(k-1)}{SSW_n/(N-k)} \tag{1.2.1}$$

The distribution of the sequence $\{F_n\}$ depends only upon the noncentrality parameter λ. Applying Cox's theorem, a sequential test for discriminating between the hypotheses

$$H_0: \ \lambda = \lambda_0 \quad \text{vs.} \quad H_1: \ \lambda = \lambda_1 \ , \ \lambda_1 > \lambda_0 \tag{1.2.2}$$

for a given α and β is specified by the decision rule

$$\textit{Accept } H_0 \text{ if } \quad \frac{f(F_n|\lambda_1)}{f(F_n|\lambda_0)} \leq \frac{\beta}{1-\alpha}$$

$$\text{Reject } H_0 \text{ if } \quad \frac{f(F_n|\lambda_1)}{f(F_n|\lambda_0)} \geq \frac{1-\beta}{\alpha}$$

otherwise continue to the next stage. (1.2.3)

An equivalent test was derived by Hoel (1955) using Wald's method of weight functions. The weight function Hoel employed was a generalization of that used for Wald's sequential t-test.

The same sequential test (i.e., the test statistic of (1.2.1) and decision rule (1.2.3) of the hypotheses (1.2.2) has also been by Hall, Wijsman and Ghosh (1965). Their derivation involved applying the principal of invariance. They showed that test statistic of equation (1.2.1) is unchanged by any of the following transformations:

(i) $X'_{ijn} = CX_{ijn} \quad C > 0$

(ii) $X'_{ijn} = X_{ijn} + C$

(iii) an orthogonal transformation

Also, they were able to prove that the sequential test was UMP for testing the hypotheses $H_0: \lambda \leq \lambda_0$ vs. $H_1: \lambda \geq \lambda_1$, by showing that the density $f(F_n|\lambda)$ possessed a monotone likelihood ratio (Lehman (1959)).

In addition, they proved that the vector of statistics $T_n = \{\bar{X}_1, \bar{X}_2, \ldots, SSW_n\}$ was a transitive sufficient sequence. This finding is of importance in later chapters of the thesis.

As previously explained, the sequential test is carried out in stages, where at each stage a fixed number r_i, for $i = 1, \ldots, k$, of observations are taken from each group. Throughout the remainder of this thesis it will be assumed that at the first stage two observations from each group will be taken (this is so the statistic SSB_1 will not be zero on

the first stage). Each subsequent stage will result in one observation from each group being taken (i.e., $r_i = 1$ for all i). All future discussions will pertain to this particular testing situation.

As in the fixed sample test, the density of the statistic F_{nj} $(F_n|\lambda)$, is that of a noncentral F variate and is given in equation (1.1.1). Therefore, the decision rule of equation (1.2.3) requires calculating the ratio of two noncentral F densities. For specified values of α, β, λ_0 and λ_1 the decision rule can be reexpressed as:

accept H_0 if $\Lambda_n \leq \dfrac{\beta}{1-\alpha}$

reject H_0 if $\Lambda_n \geq \dfrac{1-\beta}{\alpha}$

continue otherwise

where

$$\Lambda = R(F_n) = \frac{e^{-\frac{n}{2}(\lambda_1-\lambda_0)}\, M\left[\frac{N-1}{2}, \frac{K-1}{2}, \frac{\lambda_0(K-1)F_n}{2\left(K(n-1)+(K-1)F_n\right)}\right]}{M\left[\frac{N-1}{2}, \frac{K-1}{2}, \frac{\lambda_1(K-1)F_n}{2\left(K(n-1)+(K-1)F_n\right)}\right]}$$

and $M(x,y,u)$, known as the confluent hypergeometric function is defined as

$$M(x,y,u) = \sum_{t=0}^{\infty} \frac{\Gamma(y)\ \Gamma(x+t)}{\Gamma(x)\ \Gamma(y+t)} \frac{u^t}{t!}$$

Since the above decision rule is a function of the statistic, F_n, the equations may be solved to obtain a decision rule in terms of that statistic. That is, two

critical values of the statistic may be found; F_n^A and F_n^R such that $R(F_n^A) = \beta/(1-\alpha)$ and $R(F_n^R) = (1-\beta)/$. When these critical values have been calculated for all stages, F_n^A and F_n^R, $n = 2,\ldots,m_0$; the sequential test can then be conducted by comparing the statistic, F_n, of equation (1.2.1) against these critical values. In summary, at every stage n the following decision rule is applied:

$$\begin{array}{lll} \text{accept } H_0 & \text{if} & F_n \leq F_n^A \\ \text{reject } H_0 & \text{if} & F_n \geq F_n^R \\ \text{continue} & \text{if} & F_n^A < F_n > F_n^R \end{array}$$

The test is usually performed using the somewhat simpler statistic

$$V_n = \frac{SSB_n}{SSW_n} \quad .$$

The relationship between the two statistics F_n and V_n is simply

$$\frac{(N-K)V_n}{(K-1)} = F_n \quad .$$

Conducting the test with the statistic V_n requires transforming the critical region as well $\{$e.g., $V_n^A = (K-1)F_n^A/(N-K)\}$.

Tables of the critical values have been prepared for selected values of α, β, K, λ_0 and λ_1 by Ray (1956) and B.K. Ghosh, et al. (1967). However, these tables are in terms of the test statistic $G_n = V_n/K$. Appendix B of this thesis contains a computer program which calculates the critical values of V_n; V_n^A and V_n^R, for specified values of α, β, K, λ_0, and λ.

As with all statistical tests, one important property of the test described above is the Operating Characteristic Curve. The OC curve for the above test is strictly a function of λ, and is given by

$$OC(\lambda^*) = Pr\{\text{accepting } H_0: \lambda = \lambda_0 \text{ if } \lambda = \lambda^*\}$$

Wald developed an approximation for the OC curve of a sequential probability ratio test of $f(X,\Theta_0)$ against $f(X,\Theta_1)$ provided the equation

$$E_\Theta\{[f(X,\Theta_1)/f(X,\Theta_0)]^h\} = 1$$

has a nonzero solution $h = h(\Theta)$, and the $\{X_1\}$ are i.i.d. However, since the above test is conducted on the transformed sequence $\{V_i\}$ which are not independent, Wald's approximation is not valid. Bhate (1955) developed a conjectural formula, similar to Wald's approximation for the OC curve, when the $\{X_i\}$ are not independent. Ghosh (1970) suggests that substituting the sequence $\{V_i\}$ into Bhate's formula may yield a useful approximation to the OC curve. The result of this substitution yields the following approximation to the OC curve.

If $h_i(\lambda)$ is a nonzero solution of the equation

$$\frac{f_i(V_i|V_1,\cdots,V_{i-1};\lambda_1)}{f_i(V_i|V_1,\cdots,V_{i-1};\lambda_0)}^{h_i} dF(V_i|V_1,\cdots,V_{i-1};\lambda) = 1$$

and $h_i(\lambda) \approx h(\lambda)$ for all $i \geq 1$, that is $h_i(\lambda)$ varies very little with i for a given λ, then

$$OC(\lambda) \simeq \frac{e^{Ah(\lambda)}-1}{e^{Ah(\lambda)}-e^{Bh(\lambda)}} .$$

Where

$$A \simeq \ln \frac{1-\beta}{\alpha} \qquad\qquad B \simeq \ln \frac{\beta}{1-\alpha}$$

The crucial point in the use of the conjecture lies in the verification of $h_i(\theta) \simeq h(\theta)$ for various values of i. Also it must be noted that this approximation is only valid for infinite Wald regions.

The only other alternative, to date, for obtaining the OC curve for this type of test is to employ Monte Carlo techniques.

Also of interest in a sequential test is the Average Sample Number function. For the above test the ASN function will be defined as:

$ASN(\lambda^*)$ = Expected number of stages until a decision is reached if $\lambda = \lambda^*$.

As with the OC curve, Wald's approximation to the ASN, is not valid due to the dependence of the $\{V_i\}$ sequence. No general formula (exact or approximate) for the ASN for composite hypotheses exists, but Bhate (1955) has developed

a conjectural formula along the same lines as that for the OC curve. Ray (1956) has applied Bhate's conjectural formula to the one-way fixed effect analysis of variance test, and obtained expressions for $\lambda = \lambda_0, \lambda_1$. Again, as with the OC curve this procedure is valid only for open Wald regions.

Since the regions are open, it is possible to progress through a large number of stages before a decision is reached. The number of stages will always be finite, however (Johnson, 1953). One way of assuming termination within a reasonable amount of time is to truncate the test. Truncation involves altering the Wald regions so that by some stage m_0 a decision can be made.

This thesis will be concerned with developing procedures to obtain the ASN function and OC curve for a SANOVA test with any given set of truncated regions. The following chapter contains a derivation of SANOVA for the case $k = 2$ by the Direct Method of Sequential Analysis (Aroian, 1968).

1.3 CONCLUSION

This chapter has served to introduce the SANOVA test. This thesis will pertain to obtaining the OC and ASN functions of such a test. Currently, only approximations exist, such as that of Bhate (1959), considered in this chapter. The next chapter will derive the first exact procedure for obtaining the OC and ASN of a $k=2$ SANOVA test.

CHAPTER 3

APPROXIMATE METHODS

3.0 INTRODUCTION

As discussed in Chapter I, the purpose of this thesis is to explore and develop procedures for obtaining the properties of a SANOVA test, the major properties of interest being the OC and ASN curves.

Chapter II was concerned with an exact procedure for the $k=2$ SANOVA test. The theory could be extended to the general case of a $k>2$ SANOVA test.

For the general case, the joint sufficient statistic vector would be of dimension $k+1$ and consist of the following elements:

$$X_{i_n} = \sum_{j=1}^{n} X_{ij} \qquad i = 1,\ldots,k$$

$$S_n = \sum_{i=1}^{k} \sum_{j=1}^{n} X_{i_j}^2$$

Calculating the ASN and OC curves via the direct method would now involve "carrying" a $k+1$ dimensional grid of points $f_i\ (X_{1_i}, X_{2_i}, X_{k_i}, S_i)$ from stage to stage, $i = n_1,\ldots,m_0$. The density of a given point

X_{1i}^*, $X_{2i}^*,\ldots,X_{ki}^*$, S_i^* at stage i must be evaluated by performing a k dimensional integration of the 2k+1 dimensional density, $f_{i-1}^P(X_{1i}, X_{2i},\ldots,X_{ki}, S_i, Z_1,\ldots,Z_k)$, with respect to $Z_1, Z_2,\ldots,Z_k$.

The region of integration would depend upon the point X_{1i}^*, $X_{zi}^*,\ldots,X_{ki}^*$, S_i^*, as well as the regions V_A^{i-1} and V_R^{i-1}. For the special case when no decision could be made at stage i-1, (i.e., $V_A^{i-1} < 0$ and $V_R^{i-1} > \infty$), the integration region would be that of integrating around a hypersphere. Other cases would involve integrating around pieces of hyperellipses. This process would be repeated for all points contained on the grid, yielding the density $f_i(X_{1i}, X_{2i},\ldots, X_{ki}, S_i)$.

To obtain the probabilities P_A^i, P_R^i, P_C^i would would involve performing a k+1 dimensional integration of the density $f_i(X_{1i}, X_{2i},\ldots,X_{ki}, S_i)$, the integration region in each case being that of a hyperparaboloid.

Although the direct method is theoretically possible for $k > 2$, the amount of calculations required to perform the procedure is so large as to make it impractical with the current state of computer technology. An alternative

is to use an approximate procedure for obtaining the ASN and OC curves. The purpose of this chapter is the discussion of a new multivariate normal approximation.

Since this chapter is concerned with an approximate procedure, time should be taken to discuss the current most widely used approximate technique, that of Monte Carlo simulation.

3.1 MONTE CARLO SIMULATION METHODS

Monte Carlo simulation of a SANOVA test involves performing many statistical tests of the hypotheses and evaluating its overall performance. The data for each of the tests is computer generated. Since the derivation of a SANOVA test requires that each observation X_{ij} be from a normal distribution with mean μ_i and standard deviation σ, the computer generation of data consists of generating observations from each of these normal distributions (Box and Muller (1964)).

For a given state of nature λ (i.e., μ_1, μ_2, $\ldots, \mu_k$, and σ chosen such that

$$\lambda = \sum_{i=1}^{k_1} (\mu_i - \bar{\mu})^2/\sigma^2), \text{ a large number, } N,$$

of statistical tests may be simulated. Each statistical test consists of generating vectors of observations

$$X_{(n)} = X_{1(n)}, \ldots, X_{k(n)} \quad \text{where} \quad X_{i(n)} \sim N(\mu_i, \sigma)$$

until the statistic V_n indicates acceptance or rejection of the hypothesis (i.e., $V_n \leq V_A^n$ or $V_n \geq V_R^n$).

The result of N such tests is the following tally:

f_A^i = relative frequency of accepting H_0 at stage i

f_R^i = relative frequency of rejecting H_0 at stage i.

From this tally the ASN and OC are approximated as:

$$OC \simeq \sum_{i=1}^{m_0} f_A^{\,i}$$

and

$$ASN \simeq \sum_{i=1}^{m_0} i \cdot (f_A^{\,i} + f_R^{\,i})$$

This entire process is repeated for all values of desired λ, yielding an approximate OC and ASN curve. An example of statistical simulation is given in Hahn and Shapiro (1969).

Tables 3 and 4 contain the simulation results for the two truncated SANOVA tests shown in Tables 1 and 2. Figures 15 and 16 compare the ASN and OC curves against the corresponding fixed sample test. The fixed sample OC curves were calculated exactly by the methods discussed in Appendix A.

Because Monte Carlo simulation involves random values, the results are subject to statistical fluctuations. Thus any estimate of OC or ASN will not be exact but will have an associated error band. The larger the number of trials in the simulation, the more precise will be the final answer, and we can obtain as small an error as desired by conducting sufficient trials. Theoretically as $N \to \infty$,

TABLE 1

SANOVA and ANOVA Tests # 1

FIXED SAMPLE ANOVA TEST

****

K=2.0 GROUPS

H0:LAM0 = 0.00 VS H1:LAM1 = 1.00

ALPHA = 0.10 BETA = 0.10

REQUIRED SAMPLE SIZE IS 10.0

OC FUNCTION FOR THE TEST

LAMDA	PROB OF ACCEPTING HO
0.00	0.9000
0.11	0.7278
0.22	0.5736
0.33	0.4450
0.44	0.3412
0.56	0.2594
0.67	0.1958
0.78	0.1470
0.89	0.1097
1.00	0.0815

CRITICAL VALUE OF F = 3.01

CRITICAL VALUE OF V = -0.16705

TABLE 1

(CONT.)

SEQUENTIAL ANØVA TEST
****

K= 2 MEANS

H0: LAM0 = 0 VS H1: LAM1 = 1

SEQUENTIAL TEST WITH THESE REGIØNS

N	ACCEPT	REJECT
2	***	***
3	***	24.7841
4	***	2.0189
5	.0149	1.0794
6	.0316	.7543
7	.0431	.5912
8	.0517	.4937
9	.0585	.4292
10	.0642	.3835
11	.065	.3
12	.07	.25
13	.08	.2
14	.09	.15
15	.1	.1

The Sequential Test has been Truncated at Step 15

TABLE 2

SANOVA and ANOVA Tests # 2

****

K=2.0 GROUPS

HO:LAMO = 0.00 VS HI:LAM1 = 1.00

ALPHA = 0.01 BETA = 0.01

REQUIRED SAMPLE SIZE IS 27.0

OC FUNCTION FOR THE TEST

LAMDA	PROB OF ACCEPTING HO
0.00	0.9900
0.11	0.8180
0.22	0.5793
0.33	0.3673
0.44	0.2154
0.56	0.1192
0.67	0.0631
0.78	0.0323
0.89	0.0160
1.00	0.0078

CRITICAL VALUE OF F = 7.15

CRITICAL VALUE OF V = 0.13748

TABLE 2

(CONT.)

SEQUENTIAL ANOVA TEST
****

K= 2 MEANS

LAM0 = 0.00 VS LAM1 = 1.00

A TEST WITH THESE REGIONS

N	ACCEPT	REJECT
2	*********	*********
3	*********	*********
4	*********	*********
5	*********	42.7986
6	*********	21.9807
7	*********	11.2870
8	*********	5.5505
9	*********	2.4422
10	0.0049	0.8082
11	0.0102	0.6950
12	0.0153	0.6101
13	0.0203	0.5449
14	0.0249	0.4940
15	0.0293	0.4535
16	0.0333	0.4222
17	0.0370	0.3960
18	0.0405	0.3738
19	0.0438	0.3549
20	0.0468	0.3385
21	0.0496	0.3244
22	0.0523	0.3120
23	0.0548	0.3010
24	0.0571	0.2912
25	0.0593	0.2824
26	0.0614	0.2745
27	0.0633	0.2674
28	0.0700	0.2500
29	0.0850	0.1500
30	0.1000	0.1000

The Sequential Test has been Truncated at stage 30

TABLE 3

Simulation of SANOVA Test # 1

MØNTE CARLØ SIMULATIØN FØR SANØVA

SEQUENTIAL ANØVA TEST

****

K= 2 MEANS

HO: LAMO = 0 VS H1: LAM1 = 1

SEQUENTIAL TEST WITH THESE REGIØNS

N	ACCEPT	REJECT
2	***	***
3	***	24.7841
4	***	2.0189
5	.0149	1.0794
6	.0316	.7543
7	.0431	.5912
8	.0517	.4937
9	.0585	.4292
10	.0642	.3835
11	.065	.3
12	.07	.25
13	.08	.2
14	.09	.15
15	.1	.1

TABLE 3 (Continued)

CASE NUMBER 1

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = 0 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = 0

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.999567	.433333E-3	.433333E-3
4	0	.988033	.115333E-1	.115333E-1
5	.257533	.718967	.115333E-1	.269067
6	.208967	.500167	.983333E-2	.2188
7	.130567	.362167	.743333E-2	.138
8	.868667E-1	.270333	.496667E-2	.918333E-1
9	.637667E-1	.201533	.503333E-2	.0688
10	.463667E-1	.151867	.0033	.496667E-1
11	.311333E-1	.114533	.0062	.373333E-1
12	.245333E-1	.825333E-1	.746667E-2	.032
13	.0225	.514667E-1	.856667E-2	.310667E-1
14	.0183	.219667E-1	.0112	.0295
15	.100333E-1	0	.119333E-1	.219667E-1

PROBABILITY OF ACCEPTING H0 = .900567
PROBABILITY OF REJECTING H0 = .994333E-1
AVERAGE SAMPLE NUMBER = 7.46313

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 2

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = .471405 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .111111

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.999033	.966667E-3	.966667E-3
4	0	.973867	.251667E-1	.251667E-1
5	.198433	.747433	.028	.226433
6	.157867	.565833	.237333E-1	.1816
7	.964667E-1	.4494	.199667E-1	.116433
8	.068	.362933	.184667E-1	.864667E-1
9	.486333E-1	.297933	.163667E-1	.065
10	.365667E-1	.247933	.134333E-1	.05
11	.026	.194967	.269667E-1	.529667E-1
12	.0223	.148767	.0239	.0462
13	.241333E-1	.950333E-1	.0296	.537333E-1
14	.0196	.420667E-1	.333667E-1	.529667E-1
15	.134333E-1	0	.286333E-1	.420667E-1

PROBABILITY OF ACCEPTING H0 = .711433
PROBABILITY OF REJECTING H0 = .288567
AVERAGE SAMPLE NUMBER = 8.1252

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 3

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = .666667 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .222222

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.998733	.126667E-2	.126667E-2
4	0	.958333	.0404	.0404
5	.153867	.761167	.0433	.197167
6	.111267	.609567	.403333E-1	.1516
7	.681667E-1	.5073	.0341	.102267
8	.483333E-1	.4287	.302667E-1	.0786
9	.0369	.364267	.275333E-1	.644333E-1
10	.296667E-1	.310133	.244667E-1	.541333E-1
11	.204333E-1	.246533	.431667E-1	.0636
12	.185333E-1	.185633	.423667E-1	.0609
13	.212333E-1	.1208	.0436	.648333E-1
14	.0201	.528667E-1	.478333E-1	.679333E-1
15	.0143	0	.385667E-1	.528667E-1

PROBABILITY OF ACCEPTING HO = .5428
PROBABILITY OF REJECTING HO = .4572
AVERAGE SAMPLE NUMBER = 8.54403

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 4

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = .816497 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .333333

0

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9979000E 00	.2100000E-02	.2100000E-02
4	.0000000E 01	.9431330E 00	.5476670E-01	.5476670E-01
5	.1188000E 00	.7625330E 00	.6180000E-01	.1806000E 00
6	.8440000E-01	.6225000E 00	.5563330E-01	.1400330E 00
7	.5136670E-01	.5233670E 00	.4776670E-01	.9913330E-01
8	.3720000E-01	.4425670E 00	.4360000E-01	.8080000E-01
9	.2810000E-01	.3768670E 00	.3760000E-01	.6570000E-01
10	.2183330E-01	.3204670E 00	.3456670E-01	.5640000E-01
11	.1490000E-01	.2483670E 00	.5720000E-01	.7210000E-01
12	.1403330E-01	.1840330E 00	.5030000E-01	.6433330E-01
13	.1533330E-01	.1153670E 00	.5333330E-01	.6866670E-01
14	.1536670E-01	.4673330E-01	.5326670E-01	.6863330E-01
15	.1100000E-01	.0000000E 01	.3573330E-01	.4673330E-01

PROBABILITY OF ACCEPTING HO = .412333
PROBABILITY OF REJECTING HO = .587667
AVERAGE SAMPLE NUMBER = 8.58383

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 5

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = .942809 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .444444

\0

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9980330E 00	.1966670E-02	.1966670E-02
4	.0000000E 01	.9250670E 00	.7296670E-01	.7296670E-01
5	.9206670E-01	.7541330E 00	.7886670E-01	.1709330E 00
6	.6226670E-01	.6237330E 00	.6813330E-01	.1304000E 00
7	.3813330E-01	.5212670E 00	.6433330E-01	.1024670E 00
8	.2856670E-01	.4370000E 00	.5570000E-01	.8426670E-01
9	.1926670E-01	.3699000E 00	.4783330E-01	.6710000E-01
10	.1550000E-01	.3130330E 00	.4136670E-01	.5686670E-01
11	.1110000E-01	.2360330E 00	.6590000E-01	.7700000E-01
12	.1013330E-01	.1684670E 00	.5743330E-01	.6756670E-01
13	.1146670E-01	.1004670E 00	.5653330E-01	.6800000E-01
14	.1123330E-01	.3876670E-01	.5046670E-01	.6170000E-01
15	.8333330E-02	.0000000E 01	.3043330E-01	.3876670E-01

PROBABILITY OF ACCEPTING HO = .308067
PROBABILITY OF REJECTING HO = .691933
AVERAGE SAMPLE NUMBER = 8.4859

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 6

POPULATIØN 1 MEAN = 0 SIGMA = 1
POPULATIØN 2 MEAN = 1.05409 SIGMA = 1

SIGMA ØF SIGMAS = 0

LAMBDA = .555556

SIMULATIØN SUMMARY

N	PRØB ACC	PRØB CØN	PRØB REJ	PRØB TERM
2	0	1	0	0
3	0	.996867	.313333E-2	.313333E-2
4	0	.907733	.891333E-1	.891333E-1
5	.0693	.741233	.0972	.1665
6	.0475	.6091	.846333E-1	.132133
7	.026	.508833	.742667E-1	.100267
8	.193667E-1	.4244	.650667E-1	.844333E-1
9	.0151	.352633	.566667E-1	.717667E-1
10	.0113	.293867	.474667E-1	.587667E-1
11	.743333E-2	.213867	.725667E-1	.08
12	.623333E-2	.147667	.599667E-1	.0662
13	.803333E-2	.835333E-1	.0561	.641333E-1
14	.693333E-2	.0305	.0461	.530333E-1
15	.523333E-2	0	.252667E-1	.0305

PRØBABILITY ØF ACCEPTING HO = .222433
PRØBABILITY ØF REJECTING HO = .777567
AVERAGE SAMPLE NUMBER = 8.31024

TØTAL NUMBER ØF MØNTE CARLØ TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 7

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = 1.1547 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .666667

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.997033	.296667E-2	.296667E-2
4	0	.890633	.1064	.1064
5	.530333E-1	.725367	.112233	.165267
6	.356333E-1	.590533	.0992	.134833
7	.206333E-1	.483667	.862333E-1	.106867
8	.0139	.399267	.0705	.0844
9	.103667E-1	.329567	.593333E-1	.0697
10	.0084	.271633	.495333E-1	.579333E-1
11	.443333E-2	.191133	.760667E-1	.0805
12	.466667E-2	.127733	.587333E-1	.0634
13	.533333E-2	.688333E-1	.535667E-1	.0589
14	.0052	.225333E-1	.0411	.0463
15	.366667E-2	0	.188667E-1	.225333E-1

PROBABILITY OF ACCEPTING HO = .165267
PROBABILITY OF REJECTING HO = .834733
AVERAGE SAMPLE NUMBER = 8.09793

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 8

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = 1.24722 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .777778

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.996067	.393333E-2	.393333E-2
4	0	.872067	.124	.124
5	.384667E-1	.7012	.1324	.170867
6	.263667E-1	.562367	.112467	.138833
7	.153667E-1	.4565	.0905	.105867
8	.100333E-1	.368533	.779333E-1	.879667E-1
9	.0074	.2984	.627333E-1	.701333E-1
10	.556667E-2	.239767	.530667E-1	.586333E-1
11	.326667E-2	.160367	.761333E-1	.0794
12	.0029	.1015	.559667E-1	.588667E-1
13	.003	.525667E-1	.459333E-1	.489333E-1
14	.336667E-2	.178667E-1	.313333E-1	.0347
15	.263333E-2	0	.152333E-1	.178667E-1

PROBABILITY OF ACCEPTING HO = .118367
PROBABILITY OF REJECTING HO = .881633
AVERAGE SAMPLE NUMBER = 7.8272

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 9

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = 1.33333 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = .888889

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	0	1	0	0
3	0	.995667	.433333E-2	.433333E-2
4	0	.8548	.140867	.140867
5	.0302	.672267	.152333	.182533
6	.189333E-1	.533033	.1203	.139233
7	.011	.420767	.101267	.112267
8	.793333E-2	.3301	.827333E-1	.906667E-1
9	.0053	.2593	.0655	.0708
10	.0041	.202267	.529333E-1	.570333E-1
11	.243333E-2	.128533	.0713	.737333E-1
12	.176667E-2	.779667E-1	.0488	.505667E-1
13	.183333E-2	.0393	.368333E-1	.386667E-1
14	.196667E-2	.012	.253333E-1	.0273
15	.163333E-2	0	.103667E-1	.012

PROBABILITY OF ACCEPTING HO = .0871
PROBABILITY OF REJECTING HO = .9129
AVERAGE SAMPLE NUMBER = 7.526

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

CASE NUMBER 10

POPULATION 1 MEAN = 0 SIGMA = 1
POPULATION 2 MEAN = 1.41421 SIGMA = 1

SIGMA OF SIGMAS = 0

LAMBDA = 1

SIMULATION SUMMARY

N	PROB ACC	PROB CON	PROB REJ	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9945330E 00	.5466670E-02	.5466670E-02
4	.0000000E 01	.8270670E 00	.1674670E 00	.1674670E 00
5	.2403330E-01	.6351670E 00	.1678670E 00	.1919000E 00
6	.1446670E-01	.4935670E 00	.1271330E 00	.1416000E 00
7	.7766670E-02	.3812670E 00	.1045330E 00	.1123000E 00
8	.4633330E-02	.2949330E 00	.8170000E-01	.8633330E-01
9	.4300000E-02	.2263330E 00	.6430000E-01	.6860000E-01
10	.3100000E-02	.1704330E 00	.5280000E-01	.5590000E-01
11	.1066670E-02	.1048000E 00	.6456670E-01	.6563330E-01
12	.1333330E-02	.6150000E-01	.4196670E-01	.4330000E-01
13	.1266670E-02	.2843330E-01	.3180000E-01	.3306670E-01
14	.1566670E-02	.8633330E-02	.1823330E-01	.1980000E-01
15	.8333330E-03	.0000000E 01	.7800000E-02	.8633330E-02

PROBABILITY OF ACCEPTING HO = .643667E-1
PROBABILITY OF REJECTING HO = .935633
AVERAGE SAMPLE NUMBER = 7.22667

TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 3 (Continued)

SIMULATION SUMMARY

Test #1

λ	Probability of Accepting H_0	Average Sample #
0.00	0.9006	7.4631
0.11	0.7114	8.1252
0.22	0.5428	8.5440
0.33	0.4123	8.5838
0.44	0.3081	8.4859
0.56	0.2224	8.3102
0.67	0.1653	8.0979
0.78	0.1184	7.8272
0.89	0.0871	7.5260
1.00	0.0644	7.2267

FIGURE 15

Comparison of ASN and OC Curves of SANOVA and ANOVA Tests of Table 1

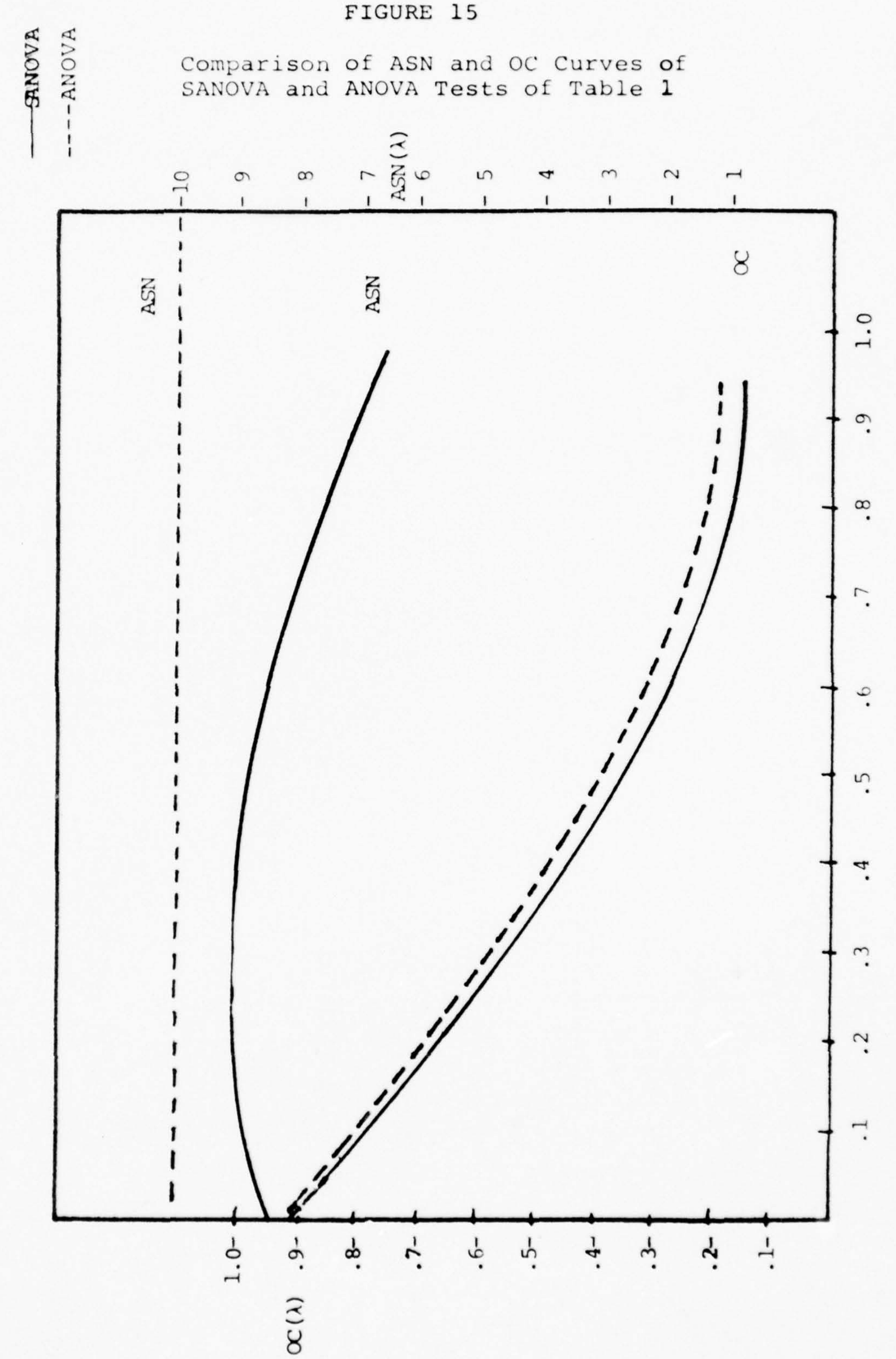


TABLE 4

Simulation of SANOVA Test # 2

MONTE CARLO SIMULATION FOR SANOVA

SEQUENTIAL TEST OF
LAM0 = 0.00 VS LAM1 = 1.00

A TEST WITH THESE REGIONS

N	ACCEPT	REJECT
2	********	********
3	********	********
4	********	********
5	********	42.7986
6	********	21.9807
7	********	11.2870
8	********	5.5505
9	********	2.4422
10	0.0049	0.8082
11	0.0102	0.6950
12	0.0153	0.6101
13	0.0203	0.5449
14	0.0249	0.4940
15	0.0293	0.4535
16	0.0333	0.4222
17	0.0370	0.3960
18	0.0405	0.3738
19	0.0438	0.3549
20	0.0468	0.3385
21	0.0496	0.3244
22	0.0523	0.3120
23	0.0548	0.3010
24	0.0571	0.2912
25	0.0593	0.2824
26	0.0614	0.2745
27	0.0633	0.2674
28	0.0700	0.2500
29	0.0850	0.1500
30	0.1000	0.1000

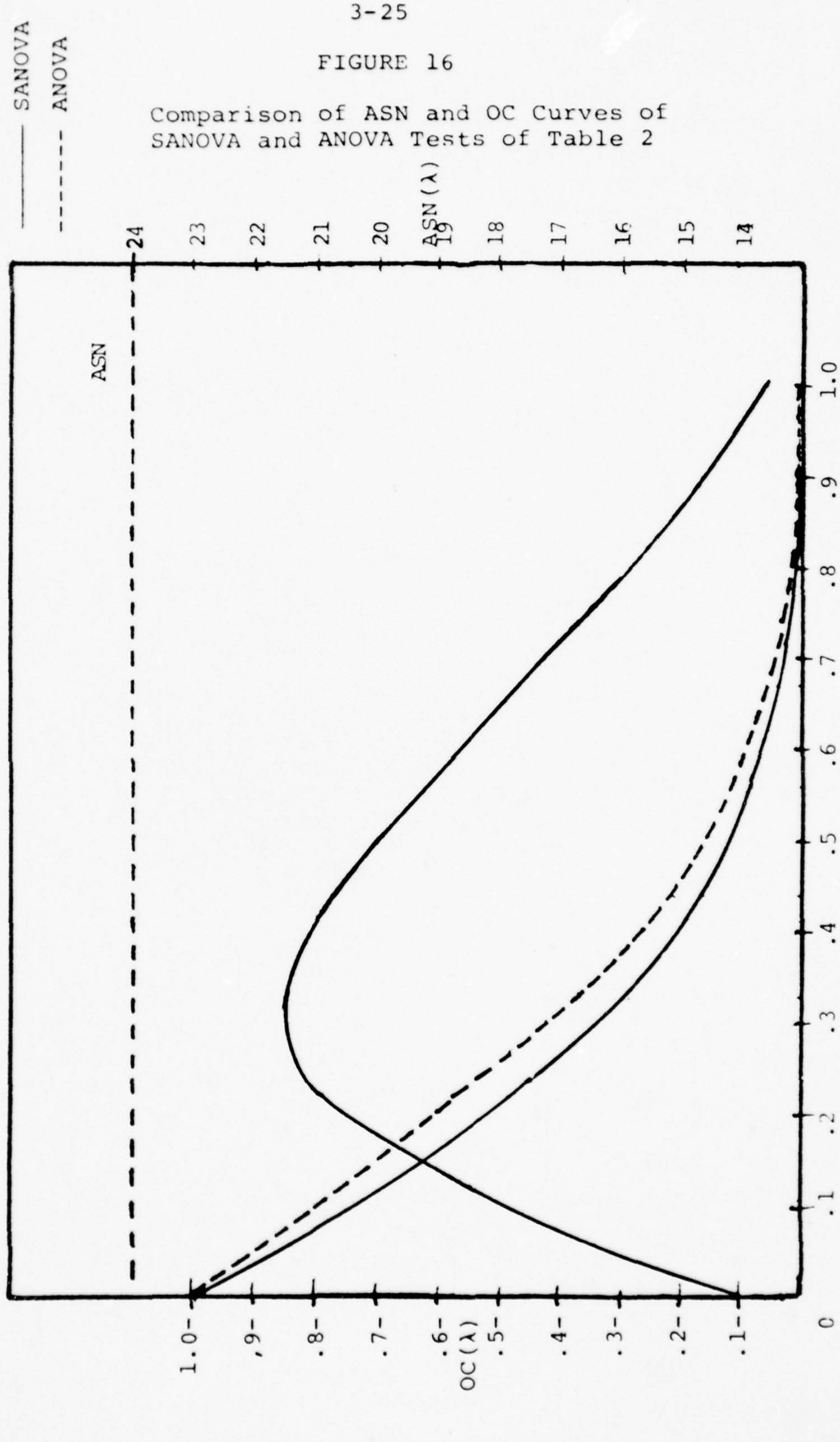
TABLE 4 (Continued)

SIMULATION SUMMARY OF SANOVA TEST #2

λ	Probability of Accepting H_0	Average Sample #
0.00	0.9826	14.3163
0.11	0.7490	18.9898
0.22	0.4912	21.2247
0.33	0.2897	21.6097
0.44	0.1660	20.8568
0.56	0.0887	19.5344
0.67	0.0449	18.1556
0.78	0.0241	16.7795
0.89	0.0122	15.5924
1.00	0.0063	14.6108

FIGURE 16

Comparison of ASN and OC Curves of
SANOVA and ANOVA Tests of Table 2



$$f_A^{\,i} \xrightarrow{P} p_A^{\,i}$$

and

$$f_R^{\,i} \xrightarrow{P} p_R^{\,i}$$

so that

$$\sum_{i=1}^{m_0} f_A^{\,i} \xrightarrow{P} \sum_{i=1}^{m_0} p^{\,i} .$$

In practice, the allowable error is generally specified, and this information can be used to determine the required number of trials, N_R. The following expression, based on the normal approximation to the binomial distribution, may be used as a rough estimate of

$$N_R \approx \frac{(.25)}{E^2} \; z^2_{1-\alpha/2}$$

where E represents the allowable error and $z_{1-\alpha/2}$ designates the $(1-\alpha/2)$ 100 percent point of a standard normal distribution.

Recently the methods of Monte Carlo importance sampling have been applied to simulations of sequential tests (Siegmund (1976)). This more sophisticated type of simulation requires smaller N for a given degree of accuracy.

3.2 MULTIVARIATE NORMAL APPROXIMATION

As previously discussed, the direct method "carries" the joint density of the sufficient statistic, $f_i(X_{1i}, X_{2i}, \ldots, X_{ki}, S_i)$ from stage to stage. At every stage this density in integrated to obtain the probabilities of acceptance, rejection, and continuation.

Consider the probability $P_A{}^i$. This quantity represents the probability of accepting H_0 at stage i. An axiom of sequential analysis requires that a decision to accept H_0 could not be made at stage i, unless all previous stages resulted in the decision to continue. Thus, this probability is a joint probability.

Since all decisions are based upon the statistic V_n, the probability $P_A{}^i$ is dependent upon the joint distribution of the statistics $V_2, V_3, \ldots, V_i$. Therefore, this probability is given by the following expression:

$$P_A{}^i = PR\left\{(V_A{}^2 < V_2 < V_R{}^2) \cap (V_A{}^3 < V_3 < V_R) \cdots \cap \right.$$

$$\left. (V_A{}^{i-1} < V_i < V_R{}^{i-1}) \cap (V_i \leq V_A{}^i)\right\}$$

One method of calculating this probability is to integrate the joint density $f(V_2, V_3, ...V_i)$, or

$$P_A^{\,i} = \int_0^{V_A^{\,i}} \int_{V_A^{\,i-1}}^{V_R^{\,i-1}} \cdots \int_{V_A^{\,2}}^{V_R^{\,2}} f(V_2, V_3, ..., V_i)\; dV_2\, dV_3 ... dV_i \tag{3.2.1}$$

Similar expressions exist for $P_R^{\,i}$ and $P_C^{\,i}$;

$$P_R^{\,i} = \int_{V_R^{\,i}}^{\infty} \int_{V_A^{\,i-1}}^{V_R^{\,i-1}} \cdots \int_{V_A^{\,2}}^{V_R^{\,2}} f(V_2, V_3, ..., V_i)\; dV_2\, dV_3 ... dV_i \tag{3.2.2}$$

$$P_C^{\,i} = \int_{V_A^{\,i}}^{V_R^{\,i}} \int_{V_A^{\,i-1}}^{V_R^{\,i-1}} \cdots \int_{V_A^{\,2}}^{V_R^{\,2}} f(V_2, V_3, ..., V_i)\; dV_2\, dV_3 ... dV_i \;. \tag{3.2.3}$$

One should also note that the identities of sequential analysis are still valid. For example,

$$P_A^{\,i} + P_R^{\,i} + P_C^{\,i} = \int_0^{\infty} \int_{V_A^{\,i-1}}^{V_R^{\,i-1}} \cdots \int_{V_A^{\,2}}^{V_R^{\,2}} f(V_2, V_3, ..., V_i\; dV_2\, dV_3 ... dV_i$$

$$= P_C^{\,i-1} \;.$$

For any stage n, the statistic V_n, where

$$V_n = \frac{T_n}{D_n} = \frac{n \sum_{i=1}^{k} \left(\overline{X}_{i(n)} - \overline{\overline{X}}_{(n)}\right)^2 / \sigma^2}{\sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \overline{X}_{i(n)})^2 / \sigma^2} \;,$$

has a distribution dependent upon λ. The numerator T_n is distributed as a noncentral χ^2 variate with non-centrality parameter $n\lambda$ and $K-1$ degrees of freedom. D_n, the denominator, is distributed as a central χ^2 variate with $K(n-1)$ degrees of freedom. Thus the statistic V_n is distributed as a constant times a noncentral F variate with noncentrality parameter $n\lambda$ and degrees freedom $K-1$, $K(n-1)$; i.e.:

$$T_n \sim \chi^2(n\lambda)_{K-1}$$

$$D_n \sim \chi^2_{K(n-1)}$$

and

$$V_n \sim \frac{K-1}{K(n-1)} F(n\lambda)_{K-1,K(n-1)} \ .$$

The statistic at stage m $(m > n)$ V_m, is correlated with the statistic at stage n; V_n, since T_n and T_m are correlated as are D_n and D_m. As shown in Appendix D:

$$\text{cov}\ (T_n, T_m) = \frac{2(K-1)n}{m} + 4n\lambda \qquad (3.2.4)$$

$$\text{cov}\ (D_n, D_m) = 2K(n-1) \qquad (3.2.5)$$

$$\text{cov}\ (T_n, D_n) = \text{cov}\ (T_m, D_m) = 0$$

$$\text{cov}\ (T_n, D_m) = \text{cov}\ (Tm, D_n) = 0 \ .$$

Therefore the joint distribution $f(V_n, V_m)$ is a type of bivariate noncentral F related distribution. Derivation of an explicit expression for the joint distribution is difficult.

The multivariate joint distribution $f(V_2, V_3, \ldots, V_i)$ is a new type of multivariate noncentral inverted Dirichlet distribution (Johnson and Kotz (1972)). Expressions for the joint density function or joint characteristic function have not been found at this time. However, it is certain that the expressions will be very complicated making the evaluation of the integrals given in equations (3.2.1) - (3.2.3) impractical. This presents a major roadblock to obtaining answers via this procedure. The best one could currently hope for would be an approximation to the density $f(V_2, V_3, \ldots, V_i)$. Several approaches to approximating the density $f(V_2, \ldots, V_i)$ are available.

One approach is the construction of a multivariate Gram-Charlier or Edgeworth series expansion. Gulberg (1920) and Meisner (1934) have given explicit formulas for a Gram-Charlier expansion of an m dimensional density function. In terms of the standardized variables,

$$X_i = (V_i - \mu_{V_i})/\sigma_{V_i}$$

$$f(X_2, X_3, \ldots, X_i) = \sum_{j_2=0}^{\infty} \sum_{j_3=0}^{\infty} \sum_{j_i=0}^{\infty} C_{j_2, j_3, \ldots, j_i} \frac{\partial^{j_2+j_3+\cdots+j_i}}{\partial X_2^{j_2} \partial X_3^{j_3} \cdots \partial X_i^{j_i}}$$

$$Z_m(X, O, R)$$

where $Z_m(X, O, R)$ is a standardized m dimensional normal density with correlation matrix R. The coefficients $C_{j_2, \ldots, j_i}$ are calculated from the mixed moments of the original distribution. Since the mixed moments can be derived by the methods given in Appendix D this approach is possible but impractical for large values of m.

Chambers (1967) has given an algorithm for the construction of Edgeworth-type expansions for a general m-variate distribution; but the technique requires the joint characteristic function, which has not yet been obtained for $f(V_2, \ldots, V_i)$.

Another approach is to transform the original statistics V_i into a new set T_i, where $T_i = g(V_2, \ldots, V_i)$, such that the joint distribution $f(T_2, \ldots, T_i)$ may be approximated by another known multivariate distribution. The procedure usually involves finding a simple univariate transformation which will approximately transform each of the marginal distributions into a known univariate distribution, and then constructing

a multivariate distribution with such marginal distributions. The most common transformation chosen is one which transforms to a normal distribution. This approach will now be considered.

First, note that since the distribution of the statistic V_n is related to an F distribution, the SANOVA test may be conducted using the statistic

$$F_n = \left(\frac{K(n-1)}{K-1}\right) V_n$$

and regions

$$F_A^{\,i} = \left(\frac{K(i-1)}{K-1}\right) V_A^{\,i}$$

$$F_R^{\,i} = \left(\frac{K(i-1)}{K-1}\right) V_R^{\,i} \tag{3.2.6}$$

The remainder of this discussion will consider the sequential test in terms of the statistic F_n.

As previously discussed the distribution of the statistic F_n is that of a noncentral F variate with noncentrality parameter $n\lambda$ and degrees of freedom $K-1$, $K(n-1)$. If $\lambda = 0$, this distribution becomes a central F variate. Consider first, possible transformations for the case $\lambda = 0$.

Several transformations have been suggested for approximately normalizing a central F variate. Aroian (1941) concluded that an excellent approximation was that suggested by Paulson (1942). The technique is based on the Wilson-Hilferty (1931) approximation to the distribution of χ^2.

The Wilson-Hilferty transformation is an approximate normalizing transformation of a χ^2 variate. If the distribution of X is χ^2 with ν degrees of freedom, then the quantity $(X/\nu)^{1/3}$ is approximately normally distributed with mean $1-(2/9\nu)$ and variance $(2/9\nu)$, i.e.

$$(X/\nu)^{1/3} \overset{A}{\sim} N\left(1 - \frac{2}{9\nu}, \sqrt{\frac{2}{9\nu}}\right). \tag{3.2.7}$$

If the distributions of T_n and D_n are each approximated in the way, then the distribution of $F_n^{1/3}$ is approximated by the distribution of the ratio of two independent normal variates. In fact, $F_n^{1/3}$ is approximately distributed as

$$\frac{1 - \dfrac{2}{9\nu_1} + U_1\sqrt{\dfrac{2}{9\nu_1}}}{1 - \dfrac{2}{9\nu_2} + U_2\sqrt{\dfrac{2}{9\nu_2}}} \tag{3.2.8}$$

where U_1, U_2 are independent normal variates and

$$\nu_1 = K-1$$

$$\nu_2 = K(n-1) .$$

The distribution of the ratio of two independent normal variates may be approximated by a method suggested by Geary (1930). This method involves the following approximation: if X_1 and X_2 are independent normal random variates and $E[X_j] = \xi_j$, $\text{var}(X_j) = \sigma_j^2$, $(j = 1,2)$ with $\xi_2 >>> \sigma_2$ then the distribution of

$$\frac{(R\xi_2 - \xi_1)}{\sqrt{R^2\sigma_2^2 + \sigma_1^2}}$$

where $R = X_1/X_2$

is approximately standard normally distributed.

Using this additional approximation for the distribution of the ratio in (3.2.6) we are led to the approximation of taking

$$Z_n = \left\{\left(1 - \frac{2}{9\nu_2}\right) F_n^{1/3} - \left(1 - \frac{2}{9\nu_1}\right)\right\}\left\{\left(\frac{2}{9\nu_2}\right)F_n^{2/3} + \left(\frac{2}{9\nu_1}\right)\right\}^{-1/2} \tag{3.2.9}$$

to have a unit normal distribution.

Each of the statistics, F_i, having marginal F distributions, may be transformed to the approximately normally distributed Z_i statistics. Thus, the original SANOVA test, using the statistic F_i and regions $F_A{}^i$, $F_R{}^i$, is approximated by a sequential test using the statistic Z_i and regions $Z_A{}^i$, $Z_R{}^i$. The probabilities $P_A{}^i$, $P_R{}^i$ and $P_C{}^i$ are then approximated by the appropriate integration of the joint density $f(Z_2,\ldots,Z_i)$.

Although each of the Z_i's is approximately normally distributed, it is not necessary that their joint distribution be approximately multinormal. An example of this has been constructed by Pierce and Dykstra (1969). In particular, they show that if

$$f(X_1,\ldots,X_m) = \left\{1 + \prod_{j=1}^{m}\left(X_j e^{-(\frac{1}{2})X_j{}^2}\right)\right\}\left\{\prod_{j=1}^{m}\left(\sqrt{\Pi}\right)e^{-(\frac{1}{2})X_j{}^2}\right\} \tag{3.2.10}$$

Then, although each subset of $X_1,\ldots,X_m$ has a joint multinormal distribution with variance-covariance matrix I, the complete set is not multinormally distributed.

Although examples like this can be constructed, many distributions are such that both the marginal and

joint distributions are normal. If in fact $f(Z_2,\ldots,Z_i)$ is not multinormal, assuming it is amounts to yet another approximation.

In summary, the multivariate normal procedure (when $\lambda = 0$) approximates the SANOVA test by a test employing the Z_n statistic and regions $Z_A{}^i$, $Z_R{}^i$, with the joint distribution $f(Z_2,\ldots,Z_n)$ being multinormal.

The joint density $f(Z_2,\ldots,Z_n)$ is given by:

$$f(Z_2,\ldots,Z_n) = f(Z) = |\Sigma|^{-1/2}(2\Pi)^{-m/2} \exp\left(-1/2\, Z'\Sigma^{-1}Z\right) \tag{3.2.11}$$

where m is the dimension (i.e., $n-1$) and Σ is the correlation matrix; i.e., $\Sigma_{ij} = \rho_{ij} = \text{corr}(Z_i, Z_j)$.

In order to specify the density, the elements of the correlation matrix Σ need be determined. To obtain the elements of this matrix exactly would require a quadrivariate integration of the joint density of T_n, D_n, T_m, D_m. This integration takes the following form:

$$E(Z_n, Z_m) = \int_{D_n}^{\infty}\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} Z_n Z_m f(T_n, D_n, T_m, D_m)\, dT_n\, dD_n\, dT_m\, dD_m \tag{3.2.12}$$

where $m > n$. (Note, the limits of integration for D_m may be justified by referring to Appendix D .)

Since T_n and D_n and T_m and D_m are independent, the joint density $f\ (T_n, D_n, T_m, D_m)$ is the product of two joint densities; i.e.

$$f\ (T_n, D_n, T_m, D_m) = f\ (T_n, T_m) \cdot f\ (D_n, D_m)\ .$$

The density $f\ (D_n, D_m)$ is a type of bivariate χ^2, and is given in Appendix D. Its form is relatively simple.

Kibble (1941) obtained the moment generating function for a bivariate gamma distribution related to $f\ (T_n, T_m)$. He obtained a Laguerre polynomial series expansion for the density.

Thus the density $f\ (T_n, D_n, T_m, D_m)$ can be expressed in a series form. Since this density is quite complicated, the integrations required to evaluate $\rho\ (Z_n, Z_m)$ can not be evaluated analytically. However, it can be evaluated numerically.

Since the Wilson-Hilferty and Geary transformations are both approximations,

$$\mu_{z_n} = \mu_{z_m} \approx 0$$

$$\sigma_{z_n} = \sigma_{z_m} \approx 1$$

and equation (3.2.12) is only an approximation to the correlation, $\rho\ (Z_n, Z_m)$. Obtaining exact values for the

correlations would involve finding exact values for μ_{Z_n}, μ_{Z_m}, σ_{Z_n}, σ_{Z_m} as well. Since the density $f\ (T_n, D_n, T_m, D_m)$ is known when $\lambda = 0$, it is possible to obtain all these quantities numerically. Thus, the distribution of Z would become a general multivariate normal with mean vector Π and variance-covariance matrix Σ. This procedure was concluded to be impractical for two reasons.

The first involves examining the accuracy of the Wilson-Hilferty approximation. Kendall and Stuart (1958) show that $(\chi^2/\nu)^{1/3}$ converges to mean $1 - (2/9\nu)$ and variance $(2/9\nu)$ at a faster rate than the rate at which the distribution approaches normality. Thus using an approximation to $\rho(Zn, Zm)$, to the same degree of accuracy as that of μ_{Z_n} and σ_{Z_n}, should have little effect on the accuracy of the overall approximation.

Second, extending the procedure for $\lambda \neq 0$ would require a much greater amount of computation. For $\lambda \neq 0$, it would be desirable to find a different transformation

$$Z_n = g_n\ (F_n, \lambda, \mu_{F_n}, \sigma_{F_n})$$

such that Z_n is approximately normally distributed. The correlations, $\rho(Z_n, Z_m)$, (which are in actuality the quantities $E\ [Z_n Z_m]$) could still be obtained as in

equation (3.2.12). However, when $\lambda \neq 0$, the joint density $f(T_n, T_m)$ is not known explicitly, rather it must be obtained by inverting the joint characteristic function, $\phi_{T_n, T_m}(t_1, t_2)$. Thus to obtain the correlations would require evaluating the following six dimensional integration:

$$\rho(Z_n, Z_m) = \int_{D_n} \int_0^\infty \int_0^\infty \int_0^\infty Z_n Z_m f(D_n, D_m) f(T_n, T_m) \, dT_n \, dD_n \, dT_m \, dD_m$$

where

$$f(T_n, T_m) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-i(t_1 T_n + t_2 T_m)\right) \phi_{T_n, T_m}(t_1, t_2) \, dt_1 \, dt_2 \tag{3.2.13}$$

and $\phi_{T_n, T_m}(t_1, t_2)$ may be obtained from the derivations of Appendix D.

Since this integration could not be done analytically it would have to be done numerically. To specify all the correlations in the correlation matrix for a SANOVA test with m^* stages at which a decision could be made, would require evaluating $1/2\,(m^{*2} - m^*)$ of the above integrals numerically.

Since exact calculation of the correlations was concluded to be impractical, an approximate procedure would be employed.

One method of approximation would involve assuming the Wilson-Hilferty $(\chi^2/\nu)^{1/3}$ was exactly normally

distributed. Thus the quantities:

$$X_1 = (T_n / \nu_{1_n})^{1/3}$$

$$X_2 = (D_n / \nu_{2_n}{}^{1/3}$$

$$X_3 = (T_m / \ {}_{1_m})^{1/3}$$

$$X_4 = (D_m / \ {}_{2_m})^{1/3} \qquad (3.2.14)$$

where $V_i = K-1$

$$V_{2_i} = K(i-1)$$

would all be assumed to be normally distributed. The correlations $\rho\ (Z_n, Z_m)$ could then be calculated as:

$$\rho\ (Z_n, Z_m) = \iiiint Z_n\, Z_m\, f\ (X_1, X_2, X_3, X_4)\ dX_1\, dX_2\, dX_3\, dX_4$$

Since X_1 and X_3 and X_2 and X_4 are correlated, the joint distribution $f\ (X_1, X_2, X_3, X_4)$ cannot be constructed as the product of the corresponding marginal distributions. Although each of the marginal distributions is normal; obviously the joint distribution is not multi-normal (since $D_m \geq D_n$).

Realizing that X_1 and X_2 and X_3 and X_4 are independent, it would not be unreasonable to construct $f\ (X_1, X_2, X_3, X_4)$ as follows:

$$f\ (X_1, X_2, X_3, X_4) = f\ (X_1, X_3) \cdot f\ (X_2, X_4),$$

where $f\ (X_1, X_2)$ is a bivariate normal distribution and $f\ (X_1, X_3)$ is a bivariate density function such that

$$\int_{-\infty}^{X_3} f\ (X_1, X_3)\ dX_1 = \frac{1}{\sqrt{2\Pi}\ \sigma_{X_3}} \exp\left\{ -\tfrac{1}{2}\left(\frac{X_3 - \mu_{X_3}}{\sigma_{X_3}}\right)^2 \right\}$$

$$\int_{X_1}^{\infty} f\ (X_1, X_3)\ dX_3 = \frac{1}{\sqrt{2\Pi}\ \sigma_{X_1}} \exp\left\{ -\tfrac{1}{2}\left(\frac{X_1 - X_1}{\sigma_{X_1}}\right)^2 \right\}.$$

Constructing the density $f\ (X_1, X_3)$ is not an easy task, and certainly its form can only complicate the integrations required to calculate $\rho(Z_n, Z_m)$. From a computational standpoint this approach does not appear to be a productive one to pursue.

This approach could be simplified by making the additional approximation that $f\ (X_1, X_2, X_3, X_4)$ is a quadrivariate normal. This would require calculating the covariances, cov (X_1, X_3) and cov (X_2, X_4), to completely specify the variance covariance matrix of the quadrivariate

normal distribution (since all the variances are assumed to be of the form $1 - (2/9\nu)$ and $\mathrm{cov}\ (X_1, X_2) = \mathrm{cov}\ (X_3, X_4) = \mathrm{cov}\ (X_1, X_4) = \mathrm{cov}\ (X_2, X_3) = 0$). These covariances are given by the following expressions:

$$\mathrm{cov}\ (X_1,X_3) = (\nu_{1n}\nu_{1m})^{-1/3}\ E\left[T_n^{1/3} T_m^{1/3}\right] - E\left[T_n^{1/3}\right] E\left[T_m^{1/3}\right]$$

$$\mathrm{cov}\ (X_2,X_4) = (\nu_{2n}\nu_{2m})^{-1/3}\ E\left[D_n^{1/3} D_m^{1/3}\right] - E\left[D_n^{1/3}\right] E\left[D_m^{1/3}\right] \tag{3.2.15}$$

It should be noted that constructing the variance-covariance matrix in this manner does not guarantee that the matrix will be positive definite. This arises from the fact that the two covariances are calculated from a joint density where $X_3 \geq X_1$; yet the multivariate normal distribution does not have this restriction. If the matrix was not positive definite, approximating $f\ (X_1, X_2, X_3, X_4)$ by a quadrivariate normal would not be appropriate.

As shown in the above equations, actual calculation of $\mathrm{cov}\ (X_1, X_3)$, $\mathrm{cov}\ (X_2, X_4)$ would require expressions for mixed fractional moments of T_n, T_m and D_n, D_m. Since the joint density $f\ (D_n, D_m)$ is explicitly known, expressions can be found for $E\ [D_n^R D_m^S]$. However, as previously discussed, for $\lambda \neq 0$ similar expressions for T_n, T_m could be obtained only through inverting the joint characteristic function.

In conclusion, the approximate procedure seems to offer no advantages over the exact method.

Another type of approximation often yielding useful results is that of statistical error propagation (Hahn and Shapiro (1967)).

Consider a complicated function of the random variates $X_1, X_2, \ldots, X_n$, say

$$W = g(X_1, X_2, \ldots, X_n) .$$

In situations where exact calculations of the moments **W** are impractical, the method of statistical error propagation may yield useful approximations.

The method consists of expanding $g(X_1, X_2, \ldots, X_n)$ about $[E(X_1), E(X_2), \ldots, E(X_n)]$, the point at which each of the variables takes on its expected value, by a multivariate Taylor series. An equation for the expected value of W is then obtained by taking expected values in the resulting expression and applying some simple algebra. Higher order moments are obtained in a similar manner.

The technique may also be used to approximate the covariance of two functions: $g(X_1, X_2, \ldots, X_n)$ and $h(X_1, X_2, \ldots, X_n)$. The covariance is given by the following expression:

$$\mathrm{cov}\left(g(X_1,X_2,\ldots,X_n),\ h(X_1,X_2,\ldots,X_n)\right)$$

$$= E\, q(X_1,X_2,\ldots,X_n) - E\, g(X_1,X_2,\ldots,X_n) \cdot E\, h(X_1,X_2,\ldots,X_n)$$

where

$$q(X_1,X_2,\ldots,X_n) = g(X_1,X_2,\ldots,X_n) \cdot h(X_1,X_2,\ldots,X_n) .$$

To find the series approximation to the covariance involves: developing Taylor series expansions for $g(X_1,\ldots,X_n)$, $h(X_1,\ldots,X_n)$, and $q(X_1,\ldots,X_n)$; taking expected values of each of these expressions; and finally subtracting the product of the first two expected value expressions from the third. The result of this yields the following expression for the covariance:

$$\mathrm{cov}\left(g(X_1,\ldots,X_n),\ h(X_1,\ldots,X_n)\right)$$

$$= \left\{ \sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} \left(\frac{\partial^* q}{\partial X_i}\right) \frac{E[(X-\mu_X)^i]}{i_1!\, i_2! \cdots i_n!} \right\}$$

$$- \left\{ \sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} \left(\frac{\partial^* h}{\partial X_i}\right)\left(\frac{\partial^* g}{\partial X_j}\right) \frac{E[(X-\mu_X)^i]\ E\,[(X-\mu_X)^j]}{i_1! \cdots i_n!\ \ j_1! \cdots j_n!} \right\}$$

where

$$E\left[(X-\mu_X)^i\right] = E\left[(X_1-\mu_{X_1})^{i_1} (X_2-\mu_{X_2})^{i_2} \cdots (X_n-\mu X_n)^{i_n}\right]$$

and

$$\frac{\partial^* q}{\partial X_i} = \frac{\partial^{i_1+i_2+\ \ +i_n} q(X_1,\cdots,X_n)}{\partial X_1^{i1}\, \partial X_2^{i1} \cdots \partial X_n^{in}} \Bigg|_{\substack{X_1=\mu_{X_1} \\ X_2=\mu_{X_2} \\ \vdots \\ X_n=\mu_{X_n}}} \qquad (3.2.16)$$

This is an exact expression but in practice only a finite number of terms are used, thus yielding an approximation. A Kth order approximation consists of retaining only those terms whose powers of the expected value sum to K or less. For example, a second order approximation is of the following form:

$$\mathrm{cov}\left(g(X_1,\ldots X_n),\ h(X_1,\ldots,X_n)\right)$$

$$\approx \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial^* h}{\partial X_i}\right)\left(\frac{\partial^* g}{\partial X_j}\right) \mathrm{cov}\ (X_i,X_j) \qquad (3.2.17)\ .$$

This method can be used to develop approximate formulas for the covariance of Z_n and Z_m. Several such approximations can be examined.

One approach consists of letting

$$Z_n = h_n(F_n) = \left\{\left(1-\frac{2}{9K(n-1)}\right)F_n^{1/3} - \left(1-\frac{2}{9(K-1)}\right)\right\}\left(\frac{2}{9K(n-1)}F^{2/3} + \frac{2}{9(K-1)}\right)^{-1/2}$$

and

$$Z_m = h_m(F_m) = \left\{\left(1-\frac{2}{9K(m-1)}\right)F_m^{1/3} - \left(1-\frac{2}{9(K-1)}\right)\right\}\left(\frac{2}{9K(m-1)}F_m^{2/3} + \frac{2}{9(K-1)}\right)^{-1/2}$$

$$(3.2.18)$$

Then a second order approximation to $\mathrm{cov}\ (Z_n,Z_m)$ is given by:

$$\mathrm{cov}\ (Z_n,Z_m) \approx \left(\frac{d^* h_n}{dF_n}\right)\left(\frac{d^* h_m}{dF_m}\right) \mathrm{cov}\ (F_n,F_m)\ .$$

The necessary partial derivatives are of the following form:

$$\frac{d^*h_i}{dF_i} = \frac{\left\{\left[\left(\frac{2}{9\nu_i}\right)\left(\frac{\nu_i}{\nu_i-2}\right)^{1/3} + \frac{2}{9\nu_i}\right] - \left(\frac{4}{81\nu_1 i}\right)\left[1 + \left(\frac{\nu_i}{\nu_i-2}\right)^{1/3}\right]\right\}}{3\left[\left(\frac{2}{9\nu_i}\right)\left(\frac{\nu_i}{\nu_i-2}\right)^{5/3} + \left(\frac{2}{9\nu_1}\right)\left(\frac{\nu}{\nu_i-2}\right)\right]^{3/2}} \qquad (3.2.19).$$

where $\nu_1 = K-1$ and $\nu_i = K(i-1)$

and

$$\operatorname{cov}(F_n, F_m) = \left(\frac{2\nu_n\nu_m}{\nu_1^2}\right)\left\{\frac{n\nu_1(\nu_m-2) + \nu_1^2 m}{m(\nu_n-2)(\nu_m-4)(\nu_m-2)}\right\} \qquad (3.2.20).$$

One difficulty with this approach is the problems encountered when higher order terms are added to the approximation. These higher order terms involve expression for the higher order moments; i.e., $E[(F_n-\mu_{R_n})^R(F_m-\mu_{F_m})^S]$, which do not exist for all n and m. Thus, a higher order approximation could only be used for a SANOVA test for which all the moments could be calculated.

An alternative approximation to $\mathrm{cov}\,(Z_n, Z_m)$ can be developed by considering Z_n and Z_m to be the following bivariate functions:

$$Z_n = h(X_1, X_2) = \frac{\left[\mu_2 X_1 - \mu_1 X_2\right]}{\left[\sigma_2{}^2 X_1{}^2 + \sigma_1{}^2 X_2{}^2\right]^{1/2}}$$

$$Z_m = g(X_3, X_4) = \frac{\left[\mu_4 X_3 - \mu_3 X_4\right]}{\left[\sigma_4{}^2 X_3{}^2 + \sigma_3{}^2 X_4{}^2\right]^{1/2}} \qquad (3.2.20)$$

with X_i being defined in (3.2.14) and

$$\mu_i = E[X_i]$$

$$\sigma_i^2 = \mathrm{var}(X_i) \ .$$

Since $\mathrm{cov}(X_1, X_2) = \mathrm{cov}(X_3, X_4) = 0$, the second order Taylor series approximation to $\mathrm{cov}(Z_n, Z_m)$ becomes

$$\mathrm{cov}\,(Z_n, Z_m) \left(\frac{\partial^* h}{\partial X_1}\right)\left(\frac{\partial^* g}{\partial X_2}\right) \mathrm{cov}\,(X_1, X_3) + \left(\frac{\partial^* h}{\partial X_2}\right)\left(\frac{\partial^* g}{\partial X_4}\right) \mathrm{cov}\,(X_2, X_4) \qquad (3.2.21).$$

where

$$\frac{\partial^* h}{\partial X_1} = \frac{\partial h}{\partial X_1} \Bigg| \begin{array}{l} x_1 = \mu \\ x_2 = \mu \end{array}$$

$$\mathrm{cov}\,(X_1, X_3) = \mathrm{cov}\left((T_n/\nu_{1n})^{1/3}, (T_m/\nu_{1m})^{1/3}\right) = (\nu_{1n}\nu_{1m})^{-1/3}\,\mathrm{cov}(T_n{}^{1/3} T_m{}^{1/3})$$

$$\mathrm{cov}\,(X_2, X_4) = \mathrm{cov}\left((D_n/\nu_{2n})^{1/3}, (D_m/\nu_{2m})^{1/3}\right) = (\nu_{2n}\nu_{2m})^{-1/3}\,\mathrm{cov}(D_n{}^{1/3} D_m{}^{1/3})$$

and

$$\mu_1 = \mu_3 \approx 1 - (2/9\nu_{1n}), \qquad \sigma_1^2 = \sigma_3^2 \approx 2/9\nu_{1n}$$

$$\mu_1 \approx 1 - (2/9\nu_{2n}) \qquad , \qquad \sigma_2^2 \approx 2/9\nu_{2n}$$

$$\mu_4 \approx 1 - (2/9\nu_{2m}) \qquad , \qquad \sigma_4^2 \approx 2/9\nu_{2m}$$

(3.2.22).

The quantities $\text{cov}\ (X_1, X_3)$ and $\text{cov}\ (X_2, X_4)$ could be calculated exactly directly from the corresponding density function. However, since the quantities σ_1^2 and μ_i are only approximations it was decided to approximate the covariances to the same degree of accuracy. Thus, each of these covariances was approximated by a second order Taylor series:

$$\text{cov}\ (X_1, X_3) \approx \left(9\nu_{1n}\nu_{1m}\right)^{-1} \text{cov}\ \left(T_n, T_m\right)$$

$$= \left(2/9\nu_{1n}\right) \left(n/m\right)$$

and

$$\text{cov}\ (X_2, X_4) \approx \left(9\nu_{2n}\nu_{2m}\right)^{-1} \text{cov}\ \left(D_n, D_m\right)$$

$$= \left(2/9\nu_{2m}\right) \qquad (3.2.23).$$

Substituting the expressions for the partial derivatives and the covariances of equation (3.2.23) into equation (3.2.21), yields the following approximation

$$\text{cov}\ (Z_n, Z_m) \approx \frac{(2/9\nu_{1n})(n/m)\mu_2\mu_4 + \mu_1\mu_3(2/9\nu_{2m})}{\left[\sigma_2{}^2\mu_1{}^2 + \sigma_1{}^2\mu_2{}^2\right]^{\frac{1}{2}}\left[\sigma_4{}^2\mu_3{}^2 + \sigma_3{}^2\mu_4{}^2\right]^{\frac{1}{2}}} \qquad (3.2.24).$$

where $\sigma_i{}^2$ and μ_i; $i = 1,\ldots,4$ are given in (3.2.22). Table 5 contains covariances calculated from this approximate formula, for several values of n and m.

The previous pages have contained discussions on several approaches for obtaining cov (Z_n, Z_m). The procedures considered were:

(1) A direct integration of the joint density $f(T_n, T_m, D_n, D_m)$, as shown in equations (3.2.12) - (3.2.13).

(2) A direct integration of the quadrivariate normal density $f(X_1, X_2, X_3, X_4)$, as shown in equations (3.2.14) - (3.2.15).

(3) A Taylor series expansion in terms of F_n and F_m, as shown in equations (3.2.18) - (3.2.20).

(4) A Taylor series expansion in terms of T_n, T_m, D_n, and D_m, as shown in equations (3.2.21) - (3.2.24).

TABLE 5

Covariances Calculated from Equation (3.2.24)

For Several Values of n and m

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.68853	0.54555	0.46188	0.40618	0.36601	0.33541	0.31117	0.29140	0.27488	0.26084	0.24871	0.23810	0.22872
3 *	0.68853	1.00000	0.80290	0.68573	0.60678	0.54930	0.50518	0.47002	0.44117	0.41698	0.39632	0.37841	0.36271	0.34879
4 *	0.54555	0.80290	1.00000	0.85672	0.75974	0.68887	0.63434	0.59076	0.55495	0.52486	0.49914	0.47682	0.45722	0.43983
5 *	0.46188	0.68573	0.85672	1.00000	0.88778	0.80564	0.74232	0.69168	0.65001	0.61498	0.58500	0.55897	0.53611	0.51582
6 *	0.40618	0.60678	0.75974	0.88778	1.00000	0.90792	0.83689	0.78002	0.73322	0.69384	0.66013	0.63085	0.60513	0.58229
7 *	0.36601	0.54930	0.68887	0.80564	0.90792	1.00000	0.92200	0.85953	0.80808	0.76479	0.72771	0.69551	0.66720	0.64207
8 *	0.33541	0.50518	0.63434	0.74232	0.83689	0.92200	1.00000	0.93238	0.87668	0.82979	0.78963	0.75474	0.72406	0.69682
9 *	0.31117	0.47002	0.59076	0.69168	0.78002	0.85953	0.93238	1.00000	0.94034	0.89011	0.84708	0.80970	0.77682	0.74763
10 *	0.29140	0.44117	0.55495	0.65001	0.73322	0.80808	0.87668	0.94034	1.00000	0.94664	0.90092	0.86119	0.82626	0.79523
11 *	0.27488	0.41698	0.52486	0.61498	0.69384	0.76479	0.82979	0.89011	0.94664	1.00000	0.95174	0.90980	0.87292	0.84016
12 *	0.26084	0.39632	0.49914	0.58500	0.66013	0.72771	0.78963	0.84708	0.90092	0.95174	1.00000	0.95596	0.91723	0.88282
13 *	0.24871	0.37841	0.47682	0.55897	0.63085	0.69551	0.75474	0.80970	0.86119	0.90980	0.95596	1.00000	0.95950	0.92353
14 *	0.23810	0.36271	0.45722	0.53611	0.60513	0.66720	0.72406	0.77682	0.82626	0.87292	0.91723	0.95950	1.00000	0.96252
15 *	0.22872	0.34879	0.43983	0.51582	0.58229	0.64207	0.69682	0.74763	0.79523	0.84016	0.88282	0.92353	0.96252	1.00000

Note : Matrix Calculated For a k=2 SANOVA Test

As mentioned throughout the discussions a procedure must be chosen which:

(1) Requires a feasible amount of computation to specify all correlations in the matrix Σ.

(2) Can easily be extended to the calculation of correlation for $\lambda \neq 0$.

(3) Is at **least** as accurate as the Wilson-Hilferty and Paulson approximations.

The approximation of procedure (4) (or that of equation (3.2.24)) does satisfy all the criteria.

The accuracy of this approximation is to the same degree as that of the Paulson approximation, since equation (3.2.24) is such that

$$\text{cov}\ (Z_n, Z_n) \approx 1$$

and the Paulson approximation yields

$$\text{cov}\ (Z_n, Z_n) = \text{var}\ (Z_n) \approx 1.$$

The addition of higher order terms onto the approximation would yield answers indicative of the inaccuracies of the Paulson approximation. The magnitude of these inaccuracies may be investigated by examining the theoretical moments of Z_n. The exact raw moments of Z_n are calculated from the following integration:

$$E\left[Z_n{}^R\right] = \int_0^\infty Z_n{}^R \, f\nu_{1n,\nu n}\left(F_n\right) dF_n \qquad (3.2.25)$$

where $f_{\nu_{1n},\nu_n}(F_n)$ is the density function of an F variate with ν_{1n},ν_n degrees of freedom. From these raw moments the variance, third central moment, and kurtosis of Z_n may be calculated. The accuracy of the Paulson approximation is determined by the rate at which

$$E\left[Z_n\right] \rightarrow 0$$

$$E\left[(Z_n - \mu_{Z_n})^2\right] \rightarrow 1$$

$$E\left[(Z_n - \mu_{Z_n})^3\right] \rightarrow 0$$

$$\frac{E\left[(Z_n - \mu_{Z_n})^3\right]}{\left[\text{var}(Z_n)\right]^2} \rightarrow 3$$

Since the raw moments of Z_n could not be obtained analytically, numerical integration was employed. Table 6 contains the exact mean, variance, third central moment, and kurtosis of Z_n, $n = 2,\ldots,30$, for the case $\nu_{1n} = k-1 = 1$ (the worst possible case).

Since the assumption of Z_n being normally distributed seemed to be more serious than the assumptions, $\mu_{Z_n} = 0$, $\sigma_{Z_n} = 1$, equation (3.2.24) was used to approximate cov (Z_n, Z_m). From this equation all elements of the matrix Σ, $\left[\Sigma_{n,m}\right] =$ cov (Z_n, Z_m), were calculated, thus

TABLE 6

Accuracy of The Paulson Approximation

INVESTIGATION OF THE ACCURACY OF THE GEARY TRANSFORMATION
USED TO TRANSFORM A CENTRAL F DISTRIBUTION TO A NORMAL
VIA METHOD OF MOMENTS

FOR K = 2
DOF1= 1
MEAN1= .777778 VAR1= .222222

STP	DOF2	MEAN2	VAR2	EX GEARY MEAN	EX GEARY VAR	EX GEARY 3RD-R-MOM	EX GEARY 4TH-R-MOM
2	2	.888889	.111111	0.020011	0.992184	0.171062	2.871900
3	4	.944444	.055556	0.000214	0.999909	0.003815	2.997300
4	6	.962963	.037037	0.000002	0.999996	0.000060	2.999950
5	8	.972222	.027778	-.000000	0.999998	0.000001	2.999990
6	10	.977778	.022222	0.000000	0.999998	-.000000	2.999990
7	12	.981481	.018519	0.000000	0.999999	0.000000	2.999990
8	14	.984127	.015873	-.000000	0.999998	-.000000	2.999990
9	16	.986111	.013889	-.000000	0.999998	-.000000	2.999990
10	18	.987654	.012346	0.000000	0.999997	0.000001	3.000000
11	20	.988889	.011111	-.000000	0.999999	-.000000	2.999990
12	22	.989899	.010101	-.000000	0.999997	-.000000	3.000000
13	24	.990741	.009259	-.000000	0.999997	0.000000	3.000000
14	26	.991453	.008547	0.000000	0.999998	0.000000	2.999990
15	28	.992064	.007937	-.000000	0.999998	-.000000	3.000000
16	30	.992593	.007407	-.000000	0.999997	-.000000	2.999990
17	32	.993056	.006944	-.000000	0.999998	0.000000	2.999990
18	34	.993464	.006536	0.000000	0.999999	0.000000	2.999990
19	36	.993827	.006173	0.000000	0.999999	-.000000	2.999990
20	38	.994152	.005848	0.000000	0.999999	0.000000	2.999990

completely specifying the approximate distribution of $Z' = [Z_2, Z_3, \ldots, Z_{m_0}]$.

In summary the multivariate normal approximation (for $\lambda=0$) involves the following approximations:

(1) Assuming the random variable $Z_n = g(F_n)$, given in equation (3.2.9) is normally distributed with mean zero and standard deviation one.

(2) Assuming the joint distribution of $Z' = \left[Z_2, Z_3, \ldots, Z_{m_0}\right]$ is multinormal with mean vector zero and variance covariance matrix Σ; its density being that of equation (3.2.11).

(3) Assuming the elements of Σ, $\left[\Sigma_{n,m}\right]$ are given by the Taylor series approximation equation (3.2.24).

To evaluate the overall accuracy of the MVN (for $\lambda = 0$) approximation Monte Carlo techniques were employed. For a given SANOVA test consisting of regions $V_A^{\,i}$, $V_R^{\,i}$, transformed regions $Z^{\,i}$, $Z^{\,i}$ were calculated, $i = 2, \ldots, m_0$. Next, the elements of $\left[\Sigma_{n,m}\right]$ were calculated for all $n = 2, \ldots, m_0$ and $m = 2, \ldots, m_0$. Then, random vectors $Z' = \left[Z_2, \ldots, Z_{m_0}\right]$ were generated from a multinormal distribution with mean vector zero and variance-covariance matrix Σ (Naylor et al (1966)). Each vector

generated was sequentially scanned until either of the following inequalities was satisfied:

$$Z_i \leq Z_A^i$$

$$Z_i \geq Z_R^i \; ; \quad i = 2,\ldots,m_0 .$$

The end result of the simulation consisted of the quantities f_A^i, f_R^i (see Section 3.1), from which OC $(\lambda = 0)_A$ and ASN $(\lambda = 0)_A$ could be calculated.

Table 7 contains the results of a simulation for the MVN approximation $(\lambda = 0)$ of the SANOVA test given in Table 1. These results should be compared with those given in Table 3 (those obtained by direct simulation of the SANOVA tests for $\lambda = 0$).

Although the frequencies f_A^i, f_R^i, $i=n_1,\ldots,m_0$, differ for each of the two simulations, the summary statistics, ASN and OC, are in remarkably good agreement. This is shown by the following comparison table (for $\lambda=0$):

	ASN	OC
MVN approx. Monte Carlo:	7.763	0.9006
SANOVA Monte Carlo:	7.463	0.9006

TABLE 7

Simulation Results For MVN Approximation To SANOVA Test#1

For $\lambda = 0$

MULTIVARIATE NROMAL APPOXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

```
.....................
.                   .
.                   .
. LAMBDA =   0      .
.                   .
.....................
```

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9989000E 00	.1100000E-02	.1100000E-02
4	.0000000E 01	.9874000E 00	.1150000E-01	.1150000E-01
5	.2687330E 00	.7081670E 00	.1050000E-01	.2792330E 00
6	.1608670E 00	.5388670E 00	.8433330E-02	.1693000E 00
7	.1182000E 00	.4144670E 00	.6200000E-02	.1244000E 00
8	.9083330E-01	.3189000E 00	.4733330E-02	.9556670E-01
9	.6986670E-01	.2455330E 00	.3500000E-02	.7336670E-01
10	.5270000E-01	.1898000E 00	.3033330E-02	.5573330E-01
11	.3743330E-01	.1467330E 00	.5633330E-02	.4306670E-01
12	.3080000E-01	.1096330E 00	.6300000E-02	.3710000E-01
13	.3050000E-01	.7050000E-01	.8633330E-02	.3913330E-01
14	.2426670E-01	.3426670E-01	.1196670E-01	.3623330E-01
15	.1640000E-01	.0000000E 01	.1786670E-01	.3426670E-01

PROBABILITY OF ACCEPTING HO = .9006
PROBABILITY OF REJECTING HO = .0994
AVERAGE SAMPLE NUMBER = 7.76317
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

It should also be noted that this particular example is one of the most difficult for the MVN approximation. Since all of the normalizing approximations improve with increasing degrees of freedom, the MVN approximation will improve for larger values of K. Thus, the MVN approximation is a reasonable approximation to consider. However, to be useful, the approach must be extended for $\lambda \neq 0$.

When $\lambda \neq 0$, the distribution of T_n becomes a noncentral χ^2 variate with noncentrality parameter $n\lambda$ and $K-1$ degrees of freedom. D_n is still distributed as a central χ^2 variate with $K(n-1)$ d.o.f. Thus, F_n becomes a noncentral F variate with noncentrality parameter $n\lambda$ and $K-1$, $K(n-1)$ degrees of freedom. A normalizing transformation must now be found for a noncentral F variate.

Finding a normalizing transformation for a noncentral F variate requires first finding a normalizing transformation for a noncentral χ^2 variate. Sankaran (1963) considered the problem of determining an index h which optimally normalizes

$$W = \chi_\nu^2(\lambda) \; / \; (\nu+\lambda)^h .$$

He showed that for

$$h = 1 - (2/3)(\nu+\lambda)(\nu+3\lambda)(\nu+2\lambda)^{-2} \qquad (3.2.26).$$

W is approximately normally distributed with mean

$$\mu_W(h) = 1 + h(h-1)(\nu+2\lambda)(\nu+\lambda)^{-2}$$

and variance

$$\sigma_W^2(h) = 2h^2(\nu+2\lambda)(\nu+\lambda)^{-2}$$

In fact, $1/3 \leq h \leq 1/2$ agreeing with the Wilson-Hilferty value $h = 1/3$, when $\lambda = 0$, and steadily increasing as λ increases. Although this normalizing transformation works well for a noncentral χ^2 variate, a dilemma arises when trying to normalize the noncentral F variate.

To apply Geary's transformation, both the numerator and denominator of F must be normalized, where

$$F = \frac{\chi^2_{\nu_1}(\lambda)/\nu_1}{\chi^2_{\nu_2}/\nu_2} .$$

Clearly, the optimal normalizing power for the denominator is always 1/3, whereas it is the h given by equation (3.2.26) for the numerator. However, for large values of ν_2, $(\chi^2_{\nu_2}/\nu_2)^P$ to any power $1/3 \leq P \leq 1/2$ will be approximately normally distributed.

Since a SANOVA test has $\nu_2 = K(n-1) >> \nu_1 = K-1$ the choice of the power so as to normalize the noncentral F variate is dominated by that which normalizes the noncentral χ^2 variate. Thus, the following quantity:

$$Z_n = \frac{\left\{\nu_{1n}F_n/\left(\nu_{1n}+n\lambda\right)\right\}^{h_n} \mu_{2_n} - \mu_{1_n}}{\left[\left\{\nu_{1n}F_n/\left(\nu_{1n}+n\lambda\right)\right\}^{2h_n}\sigma_{2n}^{2}+\sigma_{1n}^{2}\right]^{1/2}}$$

with

$$\mu_{2n} = 1 + \left\{(h_n - 1)/\nu_{2n}\right\} = 1 + \left\{h\left(h_n(h_n-1)/K(n-1)\right\}\right.$$

$$\sigma_{2n}^{2} = 2h_n^{2}/\nu_{2n} = 2h_n^{2}/K(n-1)$$

$$\nu_{2n} = K(n-1)$$

$$\mu_{1n} = 1_n + h\left(h_n - 1\right)(\nu_{1n} + 2n\lambda)\left(\nu_{1n}+n\lambda\right)^{-2}$$

$$\sigma_{1n}^{2} = 2h^{2}\left(\nu_{1n}+2n\lambda\right)\left(\nu_{1n}+n\lambda\right)^{-2}$$

$$\nu_{1n} = K-1$$

and

$$h_n = 1 - \left(2/3\right)\left(\nu_{1n}+n\lambda\right)\left(\nu_{1n}+3n\lambda\right)\left(\nu_{1n}+2n\lambda\right)^{-2}$$

will be assumed to be approximately normally distributed (Mudolkar, Chaubey, Lin (1976)). (Note this equation becomes equation (3.2.9) when $\lambda = 0$.)

As in the previous discussion, the joint distribution of $Z_2,\ldots,Z_{m_0}$ will be assumed to be multinormal with mean vector zero and variance-covariance matrix Σ, the elements of Σ, $\left[\Sigma_{n,m}\right] = \text{cov}\ (Z_n, Z_m)$.

Consider the quantities Z_n and Z_m expressed in the following form:

$$Z_n = \frac{(X_{1n}\mu_{2n} - \mu_{1n}X_{2n})}{(X_{1n}{}^2\sigma_{2n}{}^2 + X_{2n}{}^2\sigma_{1n}{}^2)^{\frac{1}{2}}}$$

and

$$Z_m = \frac{(X_{1m}\mu_{2m} - \mu_{1m}X_{2m})}{(X_{1m}{}^2\sigma_{2m}{}^2 + X_{2m}{}^2\sigma_{1m}{}^2)^{\frac{1}{2}}}$$

where

$$X_{1i} = (\nu_1 + i\lambda)^{-1}\ T_i$$

$$X_{2i} = \nu_{2i}\ D_i \tag{3.2.28}$$

with $\nu_1 = K-1$, $\nu_{2i} = K(i-1)$ and T_i, D_i are defined are defined on page 3-8.

To calculate $\text{cov}\,(Z_n, Z_m)$ the Taylor series approximation of equation (3.2.24) will again be used. This approximation becomes:

$$\text{cov}\,(Z_n, Z_m) \approx \left(\frac{\partial^* Z_n}{\partial X_{1n}}\right)\left(\frac{\partial^* Z_m}{\partial X_{1m}}\right) \text{cov}\,(X_{1n}, X_{1m})$$

$$+ \left(\frac{\partial^* Z_n}{\partial X_{2n}}\right)\left(\frac{\partial^* Z_m}{\partial X_{2m}}\right) \text{cov}\,(X_{2n}, X_{2m}) \qquad (3.2.29$$

where the partial derivative rotation is defined in equation (3.2.16).

and

$$\text{cov}\,(X_{1n}, X_{1m}) = \text{cov}\left[\left((T_n/(\nu_{1n}+n\lambda)\right)^{h_n}\left(T_m/(\nu_{1m}+m\lambda)\right)^{h_m}\right]$$

$$= (\nu_{1n}+n\lambda)^{-h_n}(\nu_{1m}+m\lambda)^{-h_m}\ \text{cov}\,(T_n^{h_n}\ T_m^{h_m})$$

and

$$\text{cov}\,(X_{2n}, X_{2m}) = \text{cov}\,\{(D_n/\nu_{2n})^{h_n}\ (D_m/\nu_{2m})^{h_m}\}$$

$$= \nu_{2n}^{-h_n}\ \nu_{2m}^{-h_m}\ \text{cov}\,(D_n^{h_n}\ D_m^{h_m})$$

$$(3.2.30).$$

Approximating these covariances by a Taylor series yields

$$\mathrm{cov}\,(X_{1n},X_{1m}) \approx \left[(\nu_1+n\lambda)(\nu_{1m}+m\lambda)\right]^{-1} h_n h_m \quad \mathrm{cov}\,(T_n,T_m)$$

$$= \left[(\nu_1+n\lambda)(\nu_{1m}+m\lambda)\right]^{-1} h_n h_m \quad \left[2\nu_{1n}(n/m)+4n\lambda\right]$$

and

$$\mathrm{cov}\,(X_{2n},X_{2m}) \approx (\nu_{2n}\nu_{2m}{}^{-1} h_n h_m \quad \mathrm{cov}\,(D_n,D_m),$$

$$= (\nu_{2n}\nu_{2m})^{-1} h_n h_m \quad (2\nu_{2n})$$

(3.2.31).

Thus, by making the necessary substitutions the approximation to $\mathrm{cov}\,(Z_n, Z_m)$ becomes:

$$\mathrm{cov}\,(Z_n,Z_m) \approx$$

$$\frac{h_n h_m \quad 2\nu_{1n}(n/m)+4n\lambda \quad (\nu_{1n}+n\lambda)(\nu_1+m\lambda)^{-1} \quad \mu_{2n}\mu_{2m} + 2\,\nu_{2m}{}^{-1} \quad \mu_{1n}\mu_{1m}h_n h_m}{\mu_{1m}{}^2\,\sigma_{2n}{}^2 + \mu_{2n}{}^2\,\sigma_{1n}{}^2 \;{}^{\frac{1}{2}} \quad \mu_{1m}{}^2\,\sigma_{2m}{}^2 + \mu_{2m}{}^2\,\sigma_{1m}{}^2 \;{}^{\frac{1}{2}}}$$

(3.2.32).

This approximation reduces to that of equation (3.2.24) whenever $\lambda = 0$. As with the case $\lambda = 0$, this approximation for $\mathrm{cov}\,(Z_n, Z_m)$ is of the same degree of accuracy as the normalizing transformation of (3.2.26), since

$$\mathrm{cov}\,(Z_n, Z_n) = \mathrm{VAR}\,(Z_n) = 1 \ .$$

Thus, the approximation of equation (3.2.32) can be used to calculate the elements of the correlation matrix, Σ, for all values of λ. Having this correlation matrix completely specifies the approximate distribution of the vector $Z' = (Z_2,\ldots,Z_i)$ given in (3.2.11).

To investigate the accuracy of the MVN approximation a Monte Carlo simulation was performed. This simulation consisted of applying the MVN approximation to the SANOVA test given in Table 1 . For a given value of λ this involved:

(1) transforming the SANOVA test regions with the transformation of (3.2.21).

(2) calculating the elements of the correlation matrix, Σ, from equation (3.2.32).

(3) generating vectors $Z' = (Z_2,\ldots,Z_{m_0})$ from a multivariate normal distribution with mean vector zero and correlation matrix Σ.

(4) sequentially scanning the elements of Z until

$$Z_i \geq Z_R^i \quad \text{or} \quad Z_i \geq Z_A^i \qquad i = 2,\ldots,m_0.$$

The results of the simulation are given in Table 8 . Table 9 contains a comparison of the ASN and OC curves for the SANOVA test and the MVN approximation. As seen from this table the MVN approximation gives remarkably accurate approximations to the OC and ASN curves of a SANOVA test. Also, it must be remembered that this particular SANOVA test (a $K=2$ test) is a worse case for the MVN approximation; more accurate approximations will be obtained for larger values of K.

The MVN approximations to OC and ASN need not be obtained by Monte Carlo simulation. An alternative is to obtain approximations to P_A^i, P_R^i by direct integration. As discussed previously, exact calculation of the probabilities requires integrating the joint density $f(F_2,\ldots,F_i)$, as shown in (3.2.3). Due to the form of the joint density $f(F_2,\ldots,F_i)$, such calculations are impractical. However, these quantities may not be approximated by integration of the much simpler density $f(Z_2,\ldots,Z_i)$.

TABLE 8

Simulation Results For MVN Approximation To SANOVA Test # 1

For $\lambda \neq 0$

LAMBDA = .111111

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1.00000	0.70557	0.55830	0.46913	0.40895	0.36536	0.33218	0.30599	0.28472	0.26705	0.25211	0.23928	0.22811	0.21830
3	0.70557	1.00000	0.79878	0.67471	0.59015	0.52853	0.48143	0.44412	0.41373	0.38844	0.36701	0.34857	0.33252	0.31839
4	0.55830	0.79878	1.00000	0.84677	0.74183	0.66513	0.60638	0.55976	0.52175	0.49008	0.46322	0.44010	0.41995	0.40221
5	0.46913	0.67471	0.84677	1.00000	0.87690	0.78675	0.71761	0.66270	0.61790	0.58054	0.54884	0.52155	0.49776	0.47680
6	0.40895	0.59015	0.74183	0.87690	1.00000	0.89758	0.81896	0.75649	0.70549	0.66295	0.62684	0.59574	0.56863	0.54473
7	0.36536	0.52853	0.66513	0.78675	0.89758	1.00000	0.91261	0.84314	0.78641	0.73908	0.69889	0.66427	0.63408	0.60748
8	0.33218	0.48143	0.60638	0.71761	0.81896	0.91261	1.00000	0.92399	0.86191	0.81010	0.76611	0.72820	0.69514	0.66600
9	0.30599	0.44412	0.55976	0.66270	0.75649	0.84314	0.92399	1.00000	0.93288	0.87686	0.82928	0.78829	0.75253	0.72101
10	0.28472	0.41373	0.52175	0.61790	0.70549	0.78641	0.86191	0.93288	1.00000	0.94000	0.88903	0.84511	0.80680	0.77303
11	0.26705	0.38844	0.49008	0.58054	0.66295	0.73908	0.81010	0.87686	0.94000	1.00000	0.94581	0.89911	0.85837	0.82246
12	0.25211	0.36701	0.46322	0.54884	0.62684	0.69889	0.76611	0.82928	0.88903	0.94581	1.00000	0.95064	0.90759	0.86963
13	0.23928	0.34857	0.44010	0.52155	0.59574	0.66427	0.72820	0.78829	0.84511	0.89911	0.95064	1.00000	0.95472	0.91480
14	0.22811	0.33252	0.41995	0.49776	0.56863	0.63408	0.69514	0.75253	0.80680	0.85837	0.90759	0.95472	1.00000	0.95820
15	0.21830	0.31839	0.40221	0.47680	0.54473	0.60748	0.66600	0.72101	0.77303	0.82246	0.86963	0.91480	0.95820	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.8562
4	-.1E 13	1.93464
5	-.778162	1.71596
6	-.501181	1.6035
7	-.34038	1.5334
8	-.225244	1.48542
9	-.134943	1.45106
10	-.591835E-1	1.42582
11	-.321377E-1	1.24175
12	.331177E-1	1.11348
13	.143439	.935299
14	.250766	.692521
15	.355891	.355891

TABLE 8

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .111111

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9979670E 00	.2033330E-02	.2033330E-02
4	.0000000E 01	.9725670E 00	.2540000E-01	.2540000E-01
5	.2152000E 00	.7343670E 00	.2300000E-01	.2382000E 00
6	.1206670E 00	.5897000E 00	.2400000E-01	.1446670E 00
7	.8916670E-01	.4785330E 00	.2200000E-01	.1111670E 00
8	.7050000E-01	.3889000E 00	.1913330E-01	.8963330E-01
9	.5386670E-01	.3178330E 00	.1720000E-01	.7106670E-01
10	.4190000E-01	.2620330E 00	.1390000E-01	.5580000E-01
11	.3113330E-01	.2088000E 00	.2210000E-01	.5323330E-01
12	.2670000E-01	.1577000E 00	.2440000E-01	.5110000E-01
13	.2710000E-01	.1025000E 00	.2810000E-01	.5520000E-01
14	.2473330E-01	.4536670E-01	.3240000E-01	.5713330E-01
15	.1746670E-01	.0000000E 01	.2790000E-01	.4536670E-01

PROBABILITY OF ACCEPTING HO = .718433
PROBABILITY OF REJECTING HO = .281567
AVERAGE SAMPLE NUMBER = 8.25627
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 8
(CONT.)

LAMBDA = .222222

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.72402	0.58533	0.50035	0.44227	0.39968	0.36690	0.34074	0.31928	0.30130	0.28596	0.27269	0.26106	0.25078
3 *	0.72402	1.00000	0.81515	0.69993	0.62043	0.56180	0.51647	0.48019	0.45035	0.42530	0.40390	0.38535	0.36909	0.35468
4 *	0.58533	0.81515	1.00000	0.86035	0.76358	0.69201	0.63658	0.59215	0.55558	0.52484	0.49856	0.47577	0.45578	0.43806
5 *	0.50035	0.69993	0.86035	1.00000	0.88813	0.80527	0.74102	0.68948	0.64703	0.61134	0.58081	0.55433	0.53109	0.51049
6 *	0.44227	0.62043	0.76358	0.88813	1.00000	0.90696	0.83478	0.77684	0.72911	0.68896	0.65461	0.62481	0.59865	0.57547
7 *	0.39968	0.56180	0.69201	0.80527	0.90696	1.00000	0.92054	0.85674	0.80417	0.75993	0.72209	0.68925	0.66042	0.63486
8 *	0.36690	0.51647	0.63658	0.74102	0.83478	0.92054	1.00000	0.93076	0.87370	0.82568	0.78459	0.74893	0.71763	0.68987
9 *	0.34074	0.48019	0.59215	0.68948	0.77684	0.85674	0.93076	1.00000	0.93873	0.88717	0.84304	0.80474	0.77112	0.74131
10 *	0.31928	0.45035	0.55558	0.64703	0.72911	0.80417	0.87370	0.93873	1.00000	0.94510	0.89810	0.85733	0.82152	0.78977
11 *	0.30130	0.42530	0.52484	0.61134	0.68896	0.75993	0.82568	0.88717	0.94510	1.00000	0.95029	0.90716	0.86928	0.83569
12 *	0.28596	0.40390	0.49856	0.58081	0.65461	0.72209	0.78459	0.84304	0.89810	0.95029	1.00000	0.95462	0.91477	0.87943
13 *	0.27269	0.38535	0.47577	0.55433	0.62481	0.68925	0.74893	0.80474	0.85733	0.90716	0.95462	1.00000	0.95826	0.92125
14 *	0.26106	0.36909	0.45578	0.53109	0.59865	0.66042	0.71763	0.77112	0.82152	0.86928	0.91477	0.95826	1.00000	0.96138
15 *	0.25078	0.35468	0.43806	0.51049	0.57547	0.63486	0.68987	0.74131	0.78977	0.83569	0.87943	0.92125	0.96138	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.70035
4	-.1E 13	1.73508
5	-.962274	1.46838
6	-.754892	1.3163
7	-.643752	1.21207
8	-.569185	1.13383
9	-.513648	1.07216
10	-.468524	1.02178
11	-.468	.809829
12	-.428112	.656816
13	-.341653	.454296
14	-.255863	.187927
15	-.170337	.170337

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .222222

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9962670E 00	.3733330E-02	.3733330E-02
4	.0000000E 01	.9558330E 00	.4043330E-01	.4043330E-01
5	.1683740E 00	.7461000E 00	.4136670E-01	.2097410E 00
6	.8300400E-01	.6243000E 00	.3880000E-01	.1218040E 00
7	.6363630E-01	.5267000E 00	.3396670E-01	.9760300E-01
8	.4956900E-01	.4467670E 00	.3036670E-01	.7993570E-01
9	.4156850E-01	.3784330E 00	.2676670E-01	.6833510E-01
10	.3300140E-01	.3227670E 00	.2266670E-01	.5566810E-01
11	.2400100E-01	.2582330E 00	.4053330E-01	.6453440E-01
12	.2310090E-01	.1976000E 00	.3753330E-01	.6063420E-01
13	.2406760E-01	.1313670E 00	.4216670E-01	.6623420E-01
14	.2333420E-01	.5833330E-01	.4970000E-01	.7303420E-01
15	.1813390E-01	.0000000E 01	.4020000E-01	.5833390E-01

PROBABILITY OF ACCEPTING HO = .551767
PROBABILITY OF REJECTING HO = .448233
AVERAGE SAMPLE NUMBER = 8.64288
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

LAMBDA = .333333

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1.00000	0.73348	0.59945	0.51675	0.45977	0.41769	0.38507	0.35889	0.33730	0.31911	0.30354	0.29001	0.27812	0.26756
3	0.73348	1.00000	0.82311	0.71224	0.63521	0.57800	0.53350	0.49768	0.46808	0.44310	0.42168	0.40305	0.38665	0.37209
4	0.59945	0.82311	1.00000	0.86666	0.77368	0.70448	0.65056	0.60710	0.57115	0.54081	0.51476	0.49210	0.47215	0.45442
5	0.51675	0.71224	0.86666	1.00000	0.89317	0.81356	0.75148	0.70142	0.65999	0.62500	0.59496	0.56881	0.54579	0.52533
6	0.45977	0.63521	0.77368	0.89317	1.00000	0.91106	0.84166	0.78568	0.73934	0.70019	0.66658	0.63732	0.61155	0.58864
7	0.41769	0.57800	0.70448	0.81356	0.91106	1.00000	0.92392	0.86253	0.81170	0.76876	0.73188	0.69977	0.67150	0.64636
8	0.38507	0.53350	0.65056	0.75148	0.84166	0.92392	1.00000	0.93360	0.87862	0.83216	0.79226	0.75752	0.72693	0.69973
9	0.35889	0.49768	0.60710	0.70142	0.78568	0.86253	0.93360	1.00000	0.94114	0.89140	0.84867	0.81147	0.77871	0.74957
10	0.33730	0.46808	0.57115	0.65999	0.73934	0.81170	0.87862	0.94114	1.00000	0.94716	0.90178	0.86226	0.82745	0.79650
11	0.31911	0.44310	0.54081	0.62500	0.70019	0.76876	0.83216	0.89140	0.94716	1.00000	0.95209	0.91037	0.87363	0.84096
12	0.30354	0.42168	0.51476	0.59496	0.66658	0.73188	0.79226	0.84867	0.90178	0.95209	1.00000	0.95619	0.91761	0.88329
13	0.29001	0.40305	0.49210	0.56881	0.63732	0.69977	0.75752	0.81147	0.86226	0.91037	0.95619	1.00000	0.95965	0.92377
14	0.27812	0.38665	0.47215	0.54579	0.61155	0.67150	0.72693	0.77871	0.82745	0.87363	0.91761	0.95965	1.00000	0.96261
15	0.26756	0.37209	0.45442	0.52533	0.58864	0.64636	0.69973	0.74957	0.79650	0.84096	0.88329	0.92377	0.96261	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.58679
4	-.1E 13	1.57817
5	-1.14344	1.27395
6	-.979705	1.09193
7	-.900116	.962289
8	-.851534	.861589
9	-.818521	.779589
10	-.793518	.710543
11	-.811066	.477434
12	-.788255	.305833
13	-.717737	.084924
14	-.646653	-.199449
15	-.57481	-.57481

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .333333

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9961330E 00	.3866670E-02	.3866670E-02
4	.0000000E 01	.9416330E 00	.5450000E-01	.5450000E-01
5	.1298720E 00	.7566330E 00	.5513330E-01	.1850060E 00
6	.6326940E-01	.6424330E 00	.5093330E-01	.1142030E 00
7	.4616880E-01	.5492330E 00	.4703330E-01	.9320210E-01
8	.3676830E-01	.4694670E 00	.4300000E-01	.7976830E-01
9	.2986810E-01	.4009000E 00	.3870000E-01	.6856810E-01
10	.2436780E-01	.3417000E 00	.3483330E-01	.5920110E-01
11	.1770080E-01	.2676330E 00	.5636670E-01	.7406750E-01
12	.1760080E-01	.2005000E 00	.4953330E-01	.6713410E-01
13	.1866750E-01	.1269330E 00	.5490000E-01	.7356750E-01
14	.1843410E-01	.5513330E-01	.5336670E-01	.7180080E-01
15	.1520060E-01	.0000000E 01	.3993330E-01	.5513390E-01

PROBABILITY OF ACCEPTING HO = .4179
PROBABILITY OF REJECTING HO = .5821
AVERAGE SAMPLE NUMBER = 8.74848
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

. LAMBDA = .444444 .

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.73861	0.60744	0.52617	0.46990	0.42814	0.39565	0.36946	0.34780	0.32951	0.31380	0.30012	0.28807	0.27735
3 *	0.73861	1.00000	0.82755	0.71917	0.64357	0.58720	0.54317	0.50762	0.47815	0.45322	0.43178	0.41309	0.39662	0.38197
4 *	0.60744	0.82755	1.00000	0.87018	0.77932	0.71144	0.65837	0.61545	0.57986	0.54973	0.52381	0.50121	0.48128	0.46354
5 *	0.52617	0.71917	0.87018	1.00000	0.89596	0.81815	0.75727	0.70801	0.66714	0.63254	0.60276	0.57679	0.55389	0.53349
6 *	0.46990	0.64357	0.77932	0.89596	1.00000	0.91330	0.84544	0.79052	0.74494	0.70634	0.67312	0.64414	0.61858	0.59582
7 *	0.42814	0.58720	0.71144	0.81815	0.91330	1.00000	0.92576	0.86567	0.81579	0.77355	0.73719	0.70547	0.67749	0.65258
8 *	0.39565	0.54317	0.65837	0.75727	0.84544	0.92576	1.00000	0.93513	0.88127	0.83566	0.79640	0.76214	0.73193	0.70502
9 *	0.36946	0.50762	0.61545	0.70801	0.79052	0.86567	0.93513	1.00000	0.94243	0.89366	0.85169	0.81507	0.78276	0.75399
10 *	0.34780	0.47815	0.57986	0.66714	0.74494	0.81579	0.88127	0.94243	1.00000	0.94827	0.90374	0.86488	0.83061	0.80008
11 *	0.32951	0.45322	0.54973	0.63254	0.70634	0.77355	0.83566	0.89366	0.94827	1.00000	0.95305	0.91208	0.87594	0.84375
12 *	0.31380	0.43178	0.52381	0.60276	0.67312	0.73719	0.79640	0.85169	0.90374	0.95305	1.00000	0.95702	0.91910	0.88533
13 *	0.30012	0.41309	0.50121	0.57679	0.64414	0.70547	0.76214	0.81507	0.86488	0.91208	0.95702	1.00000	0.96038	0.92510
14 *	0.28807	0.39662	0.48128	0.55389	0.61858	0.67749	0.73193	0.78276	0.83061	0.87594	0.91910	0.96038	1.00000	0.96326
15 *	0.27735	0.38197	0.46354	0.53349	0.59582	0.65258	0.70502	0.75399	0.80008	0.84375	0.88533	0.92510	0.96326	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.4999
4	-.1E 13	1.44603
5	-1.31201	1.10956
6	-1.17962	.902423
7	-1.12344	.751569
8	-1.09456	.632157
9	-1.07894	.533198
10	-1.06971	.448529
11	-1.10176	.197729
12	-1.09262	.105757E-1
13	-1.03485	-.225691
14	-.975675	-.525146
15	.915098	.915098

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .444444

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9946000E 00	.5400000E-02	.5400000E-02
4	.0000000E 01	.9247670E 00	.6983330E-01	.6983330E-01
5	.9993770E-01	.7522000E 00	.7263330E-01	.1725710E 00
6	.4533540E-01	.6378000E 00	.6906670E-01	.1144020E 00
7	.3336820E-01	.5453670E 00	.5906670E-01	.9243490E-01
8	.2896790E-01	.4622330E 00	.5416670E-01	.8313460E-01
9	.2276770E-01	.3931670E 00	.4630000E-01	.6906770E-01
10	.1816750E-01	.3357000E 00	.3930000E-01	.5746750E-01
11	.1163390E-01	.2588670E 00	.6520000E-01	.7683390E-01
12	.1143390E-01	.1884670E 00	.5896670E-01	.7040060E-01
13	.1376730E-01	.1172330E 00	.5746670E-01	.7123400E-01
14	.1286730E-01	.4736670E-01	.5700000E-01	.6986730E-01
15	.1060050E-01	.0000000E 01	.3676670E-01	.4736720E-01

PROBABILITY OF ACCEPTING HO = .308833
PROBABILITY OF REJECTING HO = .691167
AVERAGE SAMPLE NUMBER = 8.65788
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

LAMBDA = .555556

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.74146	0.61217	0.53187	0.47610	0.43459	0.40219	0.37603	0.35433	0.33598	0.32019	0.30642	0.29428	0.28347
3 *	0.74146	1.00000	0.83023	0.72343	0.64874	0.59290	0.54919	0.51380	0.48442	0.45952	0.43808	0.41936	0.40284	0.38813
4 *	0.61217	0.83023	1.00000	0.87234	0.78282	0.71577	0.66322	0.62065	0.58527	0.55528	0.52943	0.50687	0.48695	0.46920
5 *	0.53187	0.72343	0.87234	1.00000	0.89769	0.82099	0.76085	0.71210	0.67158	0.63721	0.60760	0.58173	0.55890	0.53855
6 *	0.47610	0.64874	0.78282	0.89769	1.00000	0.91469	0.84777	0.79351	0.74839	0.71013	0.67715	0.64835	0.62292	0.60025
7 *	0.43459	0.59290	0.71577	0.82099	0.91469	1.00000	0.92690	0.86761	0.81831	0.77650	0.74046	0.70898	0.68118	0.65640
8 *	0.40219	0.54919	0.66322	0.76085	0.84777	0.92690	1.00000	0.93607	0.88291	0.83781	0.79893	0.76498	0.73499	0.70826
9 *	0.37603	0.51380	0.62065	0.71210	0.79351	0.86761	0.93607	1.00000	0.94322	0.89505	0.85353	0.81727	0.78524	0.75668
10 *	0.35433	0.48442	0.58527	0.67158	0.74839	0.81831	0.88291	0.94322	1.00000	0.94894	0.90493	0.86649	0.83253	0.80226
11 *	0.33598	0.45952	0.55528	0.63721	0.71013	0.77650	0.83781	0.89505	0.94894	1.00000	0.95363	0.91312	0.87734	0.84545
12 *	0.32019	0.43808	0.52943	0.60760	0.67715	0.74046	0.79893	0.85353	0.90493	0.95363	1.00000	0.95753	0.92002	0.88657
13 *	0.30642	0.41936	0.50687	0.58173	0.64835	0.70898	0.76498	0.81727	0.86649	0.91312	0.95753	1.00000	0.96083	0.92590
14 *	0.29428	0.40284	0.48695	0.55890	0.62292	0.68118	0.73499	0.78524	0.83253	0.87734	0.92002	0.96083	1.00000	0.96365
15 *	0.28347	0.38813	0.46920	0.53855	0.60025	0.65640	0.70826	0.75668	0.80226	0.84545	0.88657	0.92590	0.96365	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.43011
4	-.1E 13	1.33019
5	-1.46802	.964896
6	-1.36026	.735634
7	-1.32309	.56615
8	-1.31053	.430307
9	-1.30951	.316477
10	-1.31366	.21809
11	-1.35819	-.482175E-1
12	-1.36079	-.249011
13	-1.31399	-.498752
14	-1.2651	-.811455
15	-1.2142	-1.2142

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .555556

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

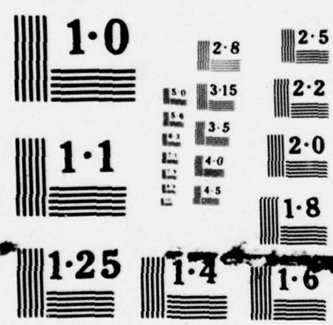
N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9926000E 00	.7400000E-02	.7400000E-02
4	.0000000E 01	.9058000E 00	.8680000E-01	.8680000E-01
5	.7127000E-01	.7450330E 00	.8950000E-01	.1607700E 00
6	.3470150E-01	.6286000E 00	.8173330E-01	.1164350E 00
7	.2586780E-01	.5259000E 00	.7683330E-01	.1027010E 00
8	.1953430E-01	.4422000E 00	.6416670E-01	.8370100E-01
9	.1526740E-01	.3709330E 00	.5600000E-01	.7126740E-01
10	.1330060E-01	.3123000E 00	.4533330E-01	.5863390E-01
11	.8633720E-02	.2330330E 00	.7063330E-01	.7926710E-01
12	.8967050E-02	.1656000E 00	.5846670E-01	.6743370E-01
13	.8667130E-02	.1001670E 00	.5676670E-01	.6543380E-01
14	.1003380E-01	.3966670E-01	.5046670E-01	.6050040E-01
15	.7833690E-02	.0000000E 01	.3183330E-01	.3966700E-01

PROBABILITY OF ACCEPTING HO = .224067
PROBABILITY OF REJECTING HO = .775933
AVERAGE SAMPLE NUMBER = 8.46191
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

AD-A072 629 UNION COLL AND UNIV SCHENECTADY NY INST OF ADMINISTR--ETC F/G 12/1
SEQUENTIAL ANALYSIS OF VARIANCE: MONTE CARLO SIMULATION, MULTIV--ETC(U)
AUG 79 R W MILLER N00014-77-C-0438
UNCLASSIFIED AES-7907 NL

2 OF 2
AD
A072629

END
DATE
FILMED
9 - 79
DDC


1.0
2.8
2.5
3.15
2.2
5.0
5.6
3.5
1.1
2.0
4.0
4.5
1.8
1.25
1.4
1.6

LAMBDA = .666667

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.74302	0.61502	0.53542	0.48002	0.43870	0.40639	0.38026	0.35855	0.34017	0.32433	0.31051	0.29831	0.28744
3 *	0.74302	1.00000	0.83195	0.72619	0.65212	0.59664	0.55315	0.51788	0.48856	0.46368	0.44224	0.42351	0.40696	0.39220
4 *	0.61502	0.83195	1.00000	0.87377	0.78513	0.71863	0.66644	0.62410	0.58886	0.55896	0.53317	0.51063	0.49072	0.47297
5 *	0.53542	0.72619	0.87377	1.00000	0.89883	0.82288	0.76324	0.71482	0.67453	0.64032	0.61081	0.58502	0.56223	0.54191
6 *	0.48002	0.65212	0.78513	0.89883	1.00000	0.91562	0.84932	0.79550	0.75069	0.71265	0.67984	0.65115	0.62580	0.60319
7 *	0.43870	0.59664	0.71863	0.82288	0.91562	1.00000	0.92765	0.86890	0.81999	0.77846	0.74263	0.71130	0.68362	0.65893
8 *	0.40639	0.55315	0.66644	0.76324	0.84932	0.92765	1.00000	0.93670	0.88399	0.83923	0.80062	0.76686	0.73702	0.71041
9 *	0.38026	0.51788	0.62410	0.71482	0.79550	0.86890	0.93670	1.00000	0.94375	0.89598	0.85476	0.81872	0.78688	0.75847
10 *	0.35855	0.48856	0.58886	0.67453	0.75069	0.81999	0.88399	0.94375	1.00000	0.94939	0.90572	0.86755	0.83381	0.80371
11 *	0.34017	0.46368	0.55896	0.64032	0.71265	0.77846	0.83923	0.89598	0.94939	1.00000	0.95401	0.91381	0.87827	0.84657
12 *	0.32433	0.44224	0.53317	0.61081	0.67984	0.74263	0.80062	0.85476	0.90572	0.95401	1.00000	0.95786	0.92062	0.88739
13 *	0.31051	0.42351	0.51063	0.58502	0.65115	0.71130	0.76686	0.81872	0.86755	0.91381	0.95786	1.00000	0.96112	0.92643
14 *	0.29831	0.40696	0.49072	0.56223	0.62580	0.68362	0.73702	0.78688	0.83381	0.87827	0.92062	0.96112	1.00000	0.96391
15 *	0.28744	0.39220	0.47297	0.54191	0.60319	0.65893	0.71041	0.75847	0.80371	0.84657	0.88739	0.92643	0.96391	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.37183
4	-.1E 13	1.226
5	-1.61309	.834361
6	-1.52588	.58505
7	-1.50499	.39876
8	-1.50661	.248102
9	-1.51842	.120841
10	-1.53441	.100459E-1
11	-1.59007	-.270226
12	-1.60314	-.483317
13	-1.5661	-.745202
14	-1.52644	-1.06987
15	-1.48427	-1.48427

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .666667

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9906330E 00	.9366670E-02	.9366670E-02
4	.0000000E 01	.8899000E 00	.1007330E 00	.1007330E 00
5	.5343570E-01	.7320330E 00	.1044330E 00	.1578690E 00
6	.2600120E-01	.6077000E 00	.9833330E-01	.1243340E 00
7	.1813420E-01	.5073330E 00	.8223330E-01	.1003680E 00
8	.1526730E-01	.4194000E 00	.7266670E-01	.8793400E-01
9	.1130050E-01	.3482000E 00	.5990000E-01	.7120050E-01
10	.8833780E-02	.2887670E 00	.5060000E-01	.5943380E-01
11	.6033620E-02	.2088670E 00	.7386670E-01	.7990030E-01
12	.5933630E-02	.1426670E 00	.6026670E-01	.6620030E-01
13	.6466960E-02	.8086670E-01	.5533330E-01	.6180030E-01
14	.6367000E-02	.3066670E-01	.4383330E-01	.5020030E-01
15	.5566930E-02	.0000000E 01	.2510000E-01	.3066690E-01

PROBABILITY OF ACCEPTING HO = .163333
PROBABILITY OF REJECTING HO = .836667
AVERAGE SAMPLE NUMBER = 8.24709
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

```
.....................
.                   .
. LAMBDA =  .777778 .
.                   .
.....................
```

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1.00000	0.74381	0.61673	0.53765	0.48254	0.44137	0.40914	0.38303	0.36134	0.34294	0.32708	0.31323	0.30100	0.29009
3	0.74381	1.00000	0.83307	0.72804	0.65440	0.59919	0.55585	0.52067	0.49139	0.46654	0.44510	0.42635	0.40978	0.39501
4	0.61673	0.83307	1.00000	0.87474	0.78671	0.72061	0.66867	0.62648	0.59135	0.56151	0.53576	0.51325	0.49334	0.47558
5	0.53765	0.72804	0.87474	1.00000	0.89963	0.82420	0.76489	0.71671	0.67658	0.64248	0.61305	0.58731	0.56455	0.54425
6	0.48254	0.65440	0.78671	0.89963	1.00000	0.91626	0.85040	0.79689	0.75230	0.71441	0.68171	0.65310	0.62781	0.60524
7	0.44137	0.59919	0.72061	0.82420	0.91626	1.00000	0.92818	0.86980	0.82116	0.77982	0.74414	0.71293	0.68533	0.66070
8	0.40914	0.55585	0.66867	0.76489	0.85040	0.92818	1.00000	0.93713	0.88475	0.84022	0.80179	0.76817	0.73844	0.71191
9	0.38303	0.52067	0.62648	0.71671	0.79689	0.86980	0.93713	1.00000	0.94411	0.89662	0.85561	0.81974	0.78802	0.75971
10	0.36134	0.49139	0.59135	0.67658	0.75230	0.82116	0.88475	0.94411	1.00000	0.94970	0.90627	0.86829	0.83469	0.80471
11	0.34294	0.46654	0.56151	0.64248	0.71441	0.77982	0.84022	0.89662	0.94970	1.00000	0.95428	0.91428	0.87892	0.84735
12	0.32708	0.44510	0.53576	0.61305	0.68171	0.74414	0.80179	0.85561	0.90627	0.95428	1.00000	0.95809	0.92103	0.88796
13	0.31323	0.42635	0.51325	0.58731	0.65310	0.71293	0.76817	0.81974	0.86829	0.91428	0.95809	1.00000	0.96132	0.92680
14	0.30100	0.40978	0.49334	0.56455	0.62781	0.68533	0.73844	0.78802	0.83469	0.87892	0.92103	0.96132	1.00000	0.96409
15	0.29009	0.39501	0.47558	0.54425	0.60524	0.66070	0.71191	0.75971	0.80471	0.84735	0.88796	0.92680	0.96409	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.32169
4	-.1E 13	1.13062
5	-1.74891	.714518
6	-1.67949	.446764
7	-1.67302	.245011
8	-1.68736	.807441E-1
9	-1.71075	-.588588E-1
10	-1.73748	-.181039
11	-1.8033	-.474129
12	-1.82592	-.69851
13	-1.79777	-.971535
14	-1.76654	-1.30722
15	-1.7324	-1.7324

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .777778

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9888670E 00	.1113330E-01	.1113330E-01
4	.0000000E 01	.8709330E 00	.1179330E 00	.1179330E 00
5	.3873510E-01	.7089670E 00	.1232330E 00	.1619680E 00
6	.1940090E-01	.5812000E 00	.1083670E 00	.1277680E 00
7	.1453390E-01	.4729000E 00	.9376670E-01	.1083010E 00
8	.1016720E-01	.3858000E 00	.7693330E-01	.8710050E-01
9	.8267040E-02	.3150670E 00	.6246670E-01	.7073370E-01
10	.6266960E-02	.2577330E 00	.5106670E-01	.5733360E-01
11	.3800200E-02	.1781000E 00	.7583330E-01	.7963350E-01
12	.3800200E-02	.1159670E 00	.5833330E-01	.6213350E-01
13	.4766880E-02	.6350000E-01	.4770000E-01	.5246690E-01
14	.4266880E-02	.2196670E-01	.3726670E-01	.4153350E-01
15	.4100190E-02	.0000000E 01	.1786670E-01	.2196690E-01

PROBABILITY OF ACCEPTING HO = .1181
PROBABILITY OF REJECTING HO = .8819
AVERAGE SAMPLE NUMBER = 7.96104
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

LAMBDA = .888889

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.74412	0.61770	0.53902	0.48414	0.44310	0.41094	0.38487	0.36319	0.34479	0.32892	0.31505	0.30280	0.29188
3 *	0.74412	1.00000	0.83381	0.72930	0.65598	0.60096	0.55773	0.52262	0.49338	0.46854	0.44710	0.42835	0.41177	0.39698
4 *	0.61770	0.83381	1.00000	0.87543	0.78783	0.72201	0.67025	0.62818	0.59313	0.56333	0.53761	0.51511	0.49521	0.47745
5 *	0.53902	0.72930	0.87543	1.00000	0.90019	0.82513	0.76608	0.71807	0.67805	0.64403	0.61465	0.58895	0.56622	0.54593
6 *	0.48414	0.65598	0.78783	0.90019	1.00000	0.91672	0.85118	0.79788	0.75345	0.71568	0.68305	0.65451	0.62926	0.60672
7 *	0.44310	0.60096	0.72201	0.82513	0.91672	1.00000	0.92855	0.87045	0.82200	0.78081	0.74523	0.71410	0.68656	0.66197
8 *	0.41094	0.55773	0.67025	0.76608	0.85118	0.92855	1.00000	0.93745	0.88529	0.84094	0.80263	0.76911	0.73946	0.71298
9 *	0.38487	0.52262	0.62818	0.71807	0.79788	0.87045	0.93745	1.00000	0.94438	0.89708	0.85622	0.82047	0.78884	0.76061
10 *	0.36319	0.49338	0.59313	0.67805	0.75345	0.82200	0.88529	0.94438	1.00000	0.94993	0.90667	0.86882	0.83533	0.80544
11 *	0.34479	0.46854	0.56333	0.64403	0.71568	0.78081	0.84094	0.89708	0.94993	1.00000	0.95447	0.91463	0.87938	0.84791
12 *	0.32892	0.44710	0.53761	0.61465	0.68305	0.74523	0.80263	0.85622	0.90667	0.95447	1.00000	0.95826	0.92134	0.88837
13 *	0.31505	0.42835	0.51511	0.58895	0.65451	0.71410	0.76911	0.82047	0.86882	0.91463	0.95826	1.00000	0.96147	0.92707
14 *	0.30280	0.41177	0.49521	0.56622	0.62926	0.68656	0.73946	0.78884	0.83533	0.87938	0.92134	0.96147	1.00000	0.96422
15 *	0.29188	0.39698	0.47745	0.54593	0.60672	0.66197	0.71298	0.76061	0.80544	0.84791	0.88837	0.92707	0.96422	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.27754
4	-.1E 13	1.04217
5	-1.87686	.603143
6	-1.82329	.318198
7	-1.8299	.102069
8	-1.85585	-.748695E-1
9	-1.88988	-.225948
10	-1.92651	-.358728
11	-2.00174	-.663734
12	-2.0332	-.898613
13	-2.0133	-1.18203
14	-1.98987	-1.52794
15	-1.96314	-1.96314

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = .888889

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9891000E 00	.1090000E-01	.1090000E-01
4	.0000000E 01	.8493330E 00	.1397670E 00	.1397670E 00
5	.2890130E-01	.6818330E 00	.1386000E 00	.1675010E 00
6	.1526730E-01	.5455000E 00	.1210670E 00	.1363340E 00
7	.1076720E-01	.4379330E 00	.9680000E-01	.1075670E 00
8	.7500340E-02	.3490670E 00	.8136670E-01	.8886700E-01
9	.5700280E-02	.2785330E 00	.6483330E-01	.7053360E-01
10	.3966880E-02	.2224000E 00	.5216670E-01	.5613350E-01
11	.2766790E-02	.1458000E 00	.7383330E-01	.7660010E-01
12	.2200130E-02	.9136670E-01	.5223330E-01	.5443350E-01
13	.2566830E-02	.4953330E-01	.3926670E-01	.4183350E-01
14	.2766810E-02	.1666670E-01	.3010000E-01	.3286680E-01
15	.1900140E-02	.0000000E 01	.1476670E-01	.1666680E-01

PROBABILITY OF ACCEPTING H0 = .0843
PROBABILITY OF REJECTING H0 = .9157
AVERAGE SAMPLE NUMBER = 7.6571
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

LAMBDA = 1

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

TRANSFORMATION YIELDS THE FOLLOWING CORRELATION MATRIX

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2 *	1.00000	0.74414	0.61820	0.53983	0.48513	0.44420	0.41211	0.38607	0.36441	0.34601	0.33015	0.31627	0.30401	0.29307
3 *	0.74414	1.00000	0.83430	0.73016	0.65707	0.60219	0.55905	0.52400	0.49479	0.46996	0.44853	0.42978	0.41319	0.39839
4 *	0.61820	0.83430	1.00000	0.87591	0.78863	0.72302	0.67139	0.62941	0.59441	0.56465	0.53895	0.51646	0.49657	0.47880
5 *	0.53983	0.73016	0.87591	1.00000	0.90060	0.82581	0.76694	0.71906	0.67912	0.64516	0.61583	0.59015	0.56744	0.54716
6 *	0.48513	0.65707	0.78863	0.90060	1.00000	0.91706	0.85175	0.79861	0.75430	0.71660	0.68404	0.65554	0.63032	0.60780
7 *	0.44420	0.60219	0.72302	0.82581	0.91706	1.00000	0.92883	0.87092	0.82262	0.78153	0.74603	0.71496	0.68746	0.66291
8 *	0.41211	0.55905	0.67139	0.76694	0.85175	0.92883	1.00000	0.93768	0.88569	0.84147	0.80326	0.76981	0.74021	0.71378
9 *	0.38607	0.52400	0.62941	0.71906	0.79861	0.87092	0.93768	1.00000	0.94457	0.89742	0.85668	0.82101	0.78945	0.76127
10 *	0.36441	0.49479	0.59441	0.67912	0.75430	0.82262	0.88569	0.94457	1.00000	0.95009	0.90696	0.86921	0.83580	0.80597
11 *	0.34601	0.46996	0.56465	0.64516	0.71660	0.78153	0.84147	0.89742	0.95009	1.00000	0.95461	0.91488	0.87972	0.84832
12 *	0.33015	0.44853	0.53895	0.61583	0.68404	0.74603	0.80326	0.85668	0.90696	0.95461	1.00000	0.95838	0.92156	0.88867
13 *	0.31627	0.42978	0.51646	0.59015	0.65554	0.71496	0.76981	0.82101	0.86921	0.91488	0.95838	1.00000	0.96158	0.92726
14 *	0.30401	0.41319	0.49657	0.56744	0.63032	0.68746	0.74021	0.78945	0.83580	0.87972	0.92156	0.96158	1.00000	0.96432
15 *	0.29307	0.39839	0.47880	0.54716	0.60780	0.66291	0.71378	0.76127	0.80597	0.84832	0.88867	0.92726	0.96432	1.00000

IN THE FOLLOWING TRANSFORMED REGION

N	ACCEPT	REJECT
2	-.1E 13	.1E 10
3	-.1E 13	2.23796
4	-.1E 13	.959342
5	-1.99808	.498665
6	-1.95889	.197559
7	-1.97753	-.320708E-1
8	-2.01426	-.220899
9	-2.05819	-.382757
10	-2.10407	-.525489
11	-2.1881	-.841677
12	-2.22782	-1.0864
13	-2.21562	-1.37954
14	-2.1995	-1.73507
15,	-2.17973	-2.17973

MULTIVARIATE NORMAL APPROXIMATION TO SEQUENTIAL ANALYSIS OF VARIANCE

LAMBDA = 1

PROBABILITIES OBTAINED VIA MONTE CARLO INTEGRATION OF MULTIVARIATE NORMAL

N	PROB ACCEPT	PROB CONTINUE	PROB REJECT	PROB TERM
2	.0000000E 01	.1000000E 01	.0000000E 01	.0000000E 01
3	.0000000E 01	.9877330E 00	.1226670E-01	.1226670E-01
4	.0000000E 01	.8340670E 00	.1536670E 00	.1536670E 00
5	.2266760E-01	.6576000E 00	.1538000E 00	.1764680E 00
6	.1103380E-01	.5142330E 00	.1323330E 00	.1433670E 00
7	.7667030E-02	.4033670E 00	.1032000E 00	.1108670E 00
8	.4866920E-02	.3174330E 00	.8106670E-01	.8593360E-01
9	.4400190E-02	.2468330E 00	.6620000E-01	.7060020E-01
10	.3333470E-02	.1915000E 00	.5200000E-01	.5533350E-01
11	.2100090E-02	.1230330E 00	.6636670E-01	.6846680E-01
12	.1533410E-02	.7463330E-01	.4686670E-01	.4840010E-01
13	.1800090E-02	.3693330E-01	.3590000E-01	.3770010E-01
14	.1333430E-02	.1133330E-01	.2426670E-01	.2560010E-01
15	.1466730E-02	.0000000E 01	.9866670E-02	.1133340E-01

PROBABILITY OF ACCEPTING HO = .0622
PROBABILITY OF REJECTING HO = .9378
AVERAGE SAMPLE NUMBER = 7.39872
TOTAL NUMBER OF MONTE CARLO TRIALS = 30000

TABLE 9

Overall Accuracy of MVN Approximation

COMPARISON OF SEQUENTIAL TEST VS MULTIVARIATE NORMAL APPROXIMATION

LAMBDA

	0.000000	0.111111	0.222222	0.333333	0.444444	0.555556	0.666667	0.777778	0.888889	1.000000
MONTE CARLO SANOVA										
OC	0.900567	0.711433	0.542800	0.412333	0.308067	0.222433	0.165267	0.118367	0.087100	0.064367
ASN	7.463130	8.125200	8.544030	8.583830	8.485900	8.310240	8.097930	7.827200	7.526000	7.226670
MVN APPROXIMATION (SANKARAN) VIA MONTE CARLO										
OC	0.900600	0.718433	0.551767	0.417900	0.308833	0.224067	0.163333	0.118100	0.084300	0.062200
ASN	7.763170	8.256270	8.642880	8.748480	8.657880	8.461910	8.247090	7.961040	7.657100	7.398720

These approximations are simply:

$$P_A{}^i \approx \int_{-\infty}^{Z_A{}^i} \int_{Z_A{}^{i-1}}^{Z_R{}^{i-1}} \cdots \int_{Z_A{}^2}^{Z_R{}^2} f\,(Z_2,\ldots,Z_i)\; dZ_2 \ldots dZ_i$$

$$P_R{}^i \approx \int_{Z_R{}^i}^{\infty} \int_{Z_A{}^{i-1}}^{Z} \cdots \int_{Z_A{}^2}^{Z} f\,(Z_2,\ldots,Z_i)\; dZ_2 \ldots dZ_i$$

(3.2.33)

where $f\,(Z_2,\ldots,Z_i)$ is the multinormal density given in (3.2.11).

Unfortunately, the integrations can not be done analytically, and thus must be obtained by numerical integration. For special forms of the correlation matrix, $\sum$, or special types of integration regions, the dimension of the integration can be reduced (Johnson and Kotz (1972)). However, none of these techniques are applicable to the integrals given in (3.2.33). Therefore, evaluation of these integrals requires an (i-1) dimensional numerical integration.

For a SANOVA test involving a large number of stages at which a decision can be made, direct integration of $f(Z_2,...,Z_i)$ becomes impractical due to the tremendous number of high dimensional numerical integrations required. Thus Monte Carlo simulation becomes the only practical method of evaluating the integrals required for the MVN approximation.

One may argue that Monte Carlo simulation of the MVN approximation seems an indirect approximate method as compared with direct Monte Carlo simulation of the SANOVA test. However, the advantages of Monte Carlo simulation of the MVN approximation are two-fold. First, for large values of K, the MVN simulation will require less time than the SANOVA simulation. Second, the MVN simulation utilizes the multinormal distribution which is currently one of the most widely investigated multivariate distributions in statistics. Thus improved simulation techniques (i.e., faster generators, importance sampling, etc.) for the MVN can easily be adapted.

3.3 CONCLUSION

This chapter of the thesis has considered approximations for the OC and ASN curves of a SANOVA test with $K > 2$ means.

The MVN approximation is a new approximation which gives remarkably accurate results. The approximation has been developed in sufficient generality to be valid for any number of means and all values of λ. The procedure requires less computation than standard Monte Carlo simulation for SANOVA tests with a large number of means. Also, the approximation will be more attractive as advances in either numerical integration or approximations for MVN probabilities are made.

CHAPTER 4

THE EFFECTS OF DEPARTURES FROM THE UNDERLYING ASSUMPTIONS IN SANOVA

4.0 INTRODUCTION

In the derivation of parametric tests it is usual to assume a form of mathematical model involving some specific probability distribution. In practical circumstances in which statistical tests are applied, little is usually known of the validity of such a model required for the procedure. The investigation of the sensitivity of the procedure to violations in the assumptions termed ("robustness," Box (1955)) has been an area of intensive research, for fixed sample procedures. Several papers, Ewens (1961), Bhattacharjee and Nagendra (1964) have investigated the sensitivity of the sequential tests to departures from assumptions.

The purpose of this chapter is to study the effects of violations of the following assumptions made in SANOVA:

(i) equality of variance of the errors

(ii) normality of the errors.

A study of this kind cannot be exhaustive, for one reason, because assumptions like this can be violated in many more ways than they can be satisfied. Therefore, the violations will be treated one at a time.

4.1 DEVIATIONS FROM HOMOGENEITY OF VARIANCE

First consider assumption (i), that of homogeneity of variances. The assumed model is that the k populations can be described by a normal distribution with mean μ_i and variance σ^2; the homogeneity of variance assumption being that each population has a common variance σ^2. Departures from this assumption occur when the populations have variances σ_i^2, not all σ_i^2 being equal. A series of papers have considered the effect of departures from this assumption in the fixed sample analysis of variance test. Box (1953), F.N. David and N.L. Johnson (1951), Horsnell (1953), Gronow (1951), Brown and Forsythe (1974), Kohr and Games (1974), Box and Andersen (1955).

Box (1953) showed that the degree of departure from the assumption could be characterized by the square of the coefficient of variation of the variances, c^2. That is to say, c^2 is the variance of the variances divided by the square of the mean of the variances $\bar{\sigma}^2$:

$$c^2 = \frac{1}{k} \sum_1 (\sigma_1^2 - \bar{\sigma}^2)/(\bar{\sigma}^2)^2 \tag{4.1.1}$$

where

$$\bar{\sigma}^2 = (\sum \sigma_i^2)/k$$

The exact distribution of the F statistic under H_0: $\lambda = 0$ when the variances were not all equal was obtained by Robbins and Pitman (1949) as an infinite series of F distributions.

Box (1953) showed that for the case of equal $n_i = n$, the distribution of the F statistic under $H_0: \lambda = 0$ could be approximated by

$$F\left\{(k-1)\varepsilon',\ k(n-1)\varepsilon\right\}$$

where $\varepsilon' = \left\{1+\frac{k-2}{k-1}\, c^2\right\}^{-1}$ and $\varepsilon = (1+c^2)^{-1}$.

Although the approximation was not of great accuracy, it did fathfully indicate the order and direction of the effects of departures.

Since ε' and ε are less than unity when the variances are not equal the significance of effects tends to be overestimated, resulting in a larger α.

The findings of Box and others concurred in demonstrating that deviations from the assumption of homogeneity of variances had very little effect on α, when the n_i's were equal and of reasonable size.

Horsnell (1953) investigated the effect of unequal group variances on the overall power of the fixed sample test. His investigation indicated that the power curve was not severely affected if one replaced the usual noncentrality parameter λ, where

$$\lambda = \frac{\sum_{i=1}^{k} n(\mu_i - \bar{\mu})^2}{\sigma^2}$$

with a modified centrality parameter λ'; where

$$\lambda' = \frac{\sum_{i=1}^{K} n(\mu_i - \bar{\mu})^2}{\bar{\sigma}^2}$$

Overall consensus is that the fixed sample analysis of variance test is robust with regard to the homogeneity of variance assumption.

As a first step toward investigating the sensitivity of SANOVA to deviations from equality of variance, a Monte Carlo simulation study was performed. For this study several sequential tests were chosen, and the OC and ASN curves were obtained via Monte Carlo simulation. Simulations were then conducted to obtain the OC and ASN at H_0 under several alternatives to the assumption of homogeneity of variances. The results of the study are summarized in Tables 10 and 11.

These results indicate that the SANOVA test is fairly robust to deviations from equality of variances assumption. Also, as in the fixed sample test, the effect can be characterized by the square of the coefficient of variation of the variances of equation (4.1.1). The magnitude of the effect may be theoretically approximated.

TABLE 10

MONTE CARLO INVESTIGATION OF HOMOGENEITY OF VARIANCE ASSUMPTION

Thesis Example: Test #2 (i.e., the test of Table 2)

Under H_0: Lambda = 0

case #	σ_1	σ_2	$\sigma^2_{\sigma_i}$	PAC	ASN	$\bar{\sigma}^2$	c^2
1	1.00	1.00	0.000	.9826	14.31	1.000	0.000
2	1.00	1.47	0.055	.9814	14.35	1.581	0.135
3	1.00	1.67	0.111	.9803	14.30	1.895	0.223
4	1.00	1.82	0.167	.9800	14.37	2.156	0.288
5	1.00	1.94	0.222	.9794	14.33	2.382	0.337
6	1.00	2.05	0.278	.9795	14.33	2.601	0.379
7	1.00	2.15	0.333	.9790	14.34	2.811	0.415
8	1.00	2.25	0.389	.9793	14.32	3.031	0.449
9	1.00	3.00	1.000	.9784	14.34	5.000	0.640
10	1.00	4.00	2.250	.9766	14.33	8.500	0.779
11	1.00	5.00	4.000	.9761	14.33	13.000	0.852
12	1.00	6.00	6.250	.9748	14.36	18.500	0.895
13	1.00	7.00	9.000	.9742	14.34	25.000	0.922

TABLE 11

MONTE CARLO INVESTIGATION OF HOMOGENEITY OF VARIANCE ASSUMPTION

A TEST WITH k=3 MEANS

All Cases for H_0: Lambda = 0

case #	σ_1	σ_2	σ_3	$\sigma^2_{\sigma_i}$	PAC	ASN	C^2
1	1	1	1	0.00	.9835	16.47	
2	1	1	3	0.89	.9526	15.49	1.058
3	1	1	4	2.00	.9446	15.14	1.39
4	1	1	5	3.56	.9384	14.96	1.58
5	1	2	4	1.56	.9597	15.74	0.85
6	1	2	5	2.89	.9522	15.44	1.14
7	1	1	2	0.22	.9688	16.02	.50

As previously discussed the SANOVA tests could be derived in terms of the statistic F_n rather than V_n, where

$$F_n = \frac{\left\{n \sum_{i=1}^{k} (\bar{X}_{i(n)} - \bar{\bar{X}}_{(n)})^2/(k-1)\right\}}{\left\{\sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{i(n)})^2/(k(n-1))\right\}}$$

The expected value of F_n when the variances are not all equal is given by

$$E[F_n] = \frac{k(n-1)\{n\hat{\lambda}+k-1\}}{(k(n-1)-2)}\, {}_2F_1(1, \tfrac{1}{2}k(n-1), k(n-1), C^2)$$

where c^2 is the square of the coefficient of variance of the population variances, and

$$\hat{\lambda} = \frac{\sum_{i=1} (\mu_i - \bar{\mu})^2}{\bar{\sigma}^2}$$

and ${}_2F_1(a,b,c,z)$ is the Gaussian hypergeometric function

$${}_2F_1(a,b,c,z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}$$

$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!}$$

so

$$E\left[F_n\right] \approx \frac{k(n-1)}{k(n-1)-2}\left[(n\hat{\lambda}+k-1)\{1+\frac{\frac{k(n-1)}{2}}{k(n-1)}c^2+\cdots\right]$$

$$\approx \frac{k(n-1)}{k(n-1)-2}\left[(n\ \ +k-1)(1+\tfrac{1}{2}c^2)\right].$$

(4.1.2)

In general, the exact distribution of F_n when the variances are not all equal is a complicated infinite series of non-central F distributions.

An approximation to the distribution of F_n is a noncentral F distribution with degrees of freedom $k-1$ and $k(n-1)$ and noncentrality parameter λ^*, chosen so $E\left[F_{n\lambda^*}(k-1,k(n-1))\right] = E\left[F_n\right]$. This results in

$$\lambda^* = \hat{\lambda}\left[1+\frac{c^2}{2}\right] + \frac{k-1}{2n^2}c^2 .$$

(4.1.3)

If this expression did not contain n, the power of the sequential test when the variances were not all equal could be approximated by $\beta(\lambda^*)$ where

$$\beta(\lambda^*) = \text{Prob}\left(\text{rejecting } H_0 \,|\, \lambda=\lambda^* \text{ when the assumptions are satisfied.}\right)$$

An alternative is to approximate the power by $\beta(\lambda')$, where

$$\lambda' = \hat{\lambda}\left\{1 + \frac{c^2}{2}\right\} + \frac{k-1}{2\left(ASN(\hat{\lambda})\right)^2 c^2} \qquad (4.1.4)$$

and

$ASN(\hat{\lambda})$ = average sample number when $\lambda = \hat{\lambda}$ when the assumptions are satisfied.

The effect of deviations from homogeniety of variance on the power of SANOVA will not only depend upon the means $\mu_1, \mu_2, \cdots, \mu_k$ and the standard deviations $\sigma_1, \sigma_2, \cdots, \sigma_k$, but also upon the following factors:

1. The combinations μ_i, σ_i
2. The number of means, k
3. The regions.

All but factor 1 are taken into consideration in the above approximation. The effects of factor 1 will be considered in the next section.

Factor 3 is taken into account in the following way: the power under a deviation from the assumption is in terms of a power when the assumption is true, which is completely determined by the regions.

Table 12 contains the results of this approximation for the simulations of Table 10. As seen from this table the approximation is fairly accurate.

TABLE 12

Departures from Normality Assumption
Test #1

case #	C^2	λ' of equation (4.1.4)	$\beta(\lambda')$ †	Observed
1	0.135	0.0003	0.9818	0.9814
2	0.223	0.0005	0.9813	0.9803
3	0.288	0.0007	0.9808	0.9800
4	0.337	0.0008	0.9806	0.9794
5	0.379	0.0009	0.9803	0.9795
6	0.415	0.0010	0.9800	0.9790
7	0.449	0.0011	0.9798	0.9793
8	0.640	0.0016	0.9786	0.9784
9	0.779	0.0019	0.9779	0.9766
10	0.852	0.0021	0.9773	0.9761
11	0.895	0.0022	0.9771	0.9748
12	0.922	0.0023	0.9766	0.9742

† Obtained by interpolation of OC curve of Table 4.

4.2 DEVIATIONS FROM NORMALITY

In considering the effects of nonnormality it is convenient to use the measures γ_1 of skewness and γ_2 of kurtosis of the distribution of the random variable X. These quantities are defined as:

$$\gamma_1 = \sigma^{-3} E[(X-\mu)^3]$$

$$\gamma_2 = \sigma^{-4} E[(X-4)^4] - 3 . \qquad (4.2.1)$$

Other commonly used measures are

$$\beta_1 = \gamma_1^2$$

for the magnitude of skewness, and

$$\beta_2 = \gamma_2 + 3$$

for kurtosis.

For a symmetrical distribution $\gamma_1 = 0$. Positive values of γ_1 indicate the distribution is "skewed to the right." Every distribution has $\gamma_2 \geq -2$, with normal distribution have $\gamma_2 = 0$. Distributions which have heavier tails and a central part more peaked than the normal have $\gamma_2 > 0$; those in which the tails are lighter and have a central part flatter, have $\gamma_2 < 0$.

Although the first four moments of a population do not determine its form entirely, the value of γ_2

and to a baser degree the value of γ_1 are the most important indicators of the extent to which nonnormality affects the usual inferences made in ANOVA.

Most of the studies of the effect of nonnormality on fixed sample ANOVA tests have:

(1) assumed that all distributions have the same skewness γ_{1i}, and the same kurtosis γ_{2i}; i.e., $\gamma_{11} = \gamma_{21} = \cdots = \gamma_{1k} = \gamma_1$ and $\gamma_{21} = \gamma_{22} = \cdots = \gamma_{2k} = \gamma_2$.

(2) dealt only with type -I errors.

Whenever each population has the same nonnormal distribution (possibly differing only with respect to location) the estimates

$$S_{T_n} = \sum_{i=1}^{k} (\bar{X}_{i(n)} - \bar{\bar{X}}_{(n)})^2 / (K-1)$$

$$S_{B_n} = \sum_{i=1}^{k} \sum_{j=1}^{n} (\bar{X}_{ij} - X_{i(n)})^2 / K(n-1)$$

still continue to provide unbiased estimates of population variance, but they are no longer independently distributed. In fact, David and Johnson (1951) showed:

$$\text{cov}\,(S_{T_n}, S_{B_n}) = \gamma_2(K^2n - nK + K + K^2)\,(n-1) / (K-1)n\,. \tag{4.2.2}$$

Since these quantities are not independent, the distribution of the ratio is obviously no longer an F distribution. Gayden (1950) approximated the distribution by an edgeworth expansion for the case $\gamma = 0$; and obtained correction terms for calculating an approximate type I error given values of γ_1 and γ_2. His calculations revealed that the effect of nonnormality diminishes rapidly in magnitude with increasing sample size, and also the effect of kurtosis was larger than that of skewness. His conclusions generally agreed with previous sampling experiments (Pearson (1931)) and later Pearson curve approximation (David and Johnson (1951)).

Box and Andersen (1955) employed permutation theory (Fisher (1935)) as a means of assessing the effects of nonnormality on the Type-I error of an ANOVA test. This resulted in the effect being represented by a modification of the degrees of freedom (ν_1^m, ν_2^m) of the F distribution. The degrees of freedom are modified in the following manner:

$$\nu_1^m = d$$

$$\nu_2^m = d(n-1)$$

where

$$d = 1 + \frac{N+1}{N-1} \frac{C_2}{N-C_2}$$

with $N = K(n-1)$

and

$$C_2 = \gamma_2 - N^{-1}\{2\gamma_4 - 3\gamma_2^{\,2} + 10\gamma_2 + 12\gamma_1^{\,2}\}$$

$$+ N^{-2}\{3\gamma_6 - 16\gamma_4\gamma_2 + 15\gamma_2^{\,3} + 36\gamma_4 + 120\gamma_3\gamma_1 - 88\gamma_1^{\,2}\gamma_2 + 66\gamma_2 + 204\gamma_1^{\,2}\}$$

(γ_i being the standardized cumulants). (4.2.3).

The modification turns out to be minor except for very small values of N.

As previously mentioned the effect of nonnormality on the probability of type-II errors has not received much attention. However, a sampling investigation for the two-tailed t-test for a single mean (Pearson (1929)) revealed that there was little effect on the power caused by non-normality.

In conclusion, it appears as though general consensus amongst statisticians is that the fixed sample ANOVA test is remarkably insensitive to nonnormality (at least the types of nonnormality considered).

Ewens (1961) considered the effect of nonnormality on the sequential test for a normal mean. He derives modified Wald approximations to the OC and ASN for such a test, when the assumption of normality is violated. His results reveal that, as in the fixed-sample tests, the sequential test for means is comparatively robust in respect to departure from assumed normality.

Evidently, the robustness of a fixed sample test to deviations from assumptions is an indication of the robustness of the analogous sequential test to the same deviations. This statement appears to hold for tests of means (Ewens (1961), Bhattacharjie and Nagendra (1964)) and variances (Wald (1947), Ewens (1961)), at least.

To investigate the sensitivity of SANOVA to deviations from the normality assumption, Monte Carlo techniques were employed. For this study several sequential tests were chosen, and the OC and ASN curves obtained via Monte Carlo simulation. This simulation being conducted under the assumption of all populations having normal distributions with mean μ_i and common variance σ^2.

A deviation from the assumption of normality involved selecting a distribution from the Johnson (1949) system of frequency curves. This system consists of three classes of distributions, S_L, S_U, and S_B, which provide one distribution corresponding to each pair of values $\sqrt{\beta}$ and β_2; i.e., there is just one appropriate distribution corresponding to each (β_1, β_2) point. Thus, a given value of β_1, β_2 specifies the type of Johnson distribution.

Thus, for a given sequential test the change in α was observed for different values of β_1, β_2. The simulation for any given pair β_1, β_2 involved generating the variates from the appropriate Johnson distribution (1949).

Tables 13 and 14 contain the OC and ASN values for several SANOVA tests with deviations from normality. These tables contain the deviational β_1, β_2 pairs together with the established α.

TABLE 13

DEPARTURES FROM NORMALITY ASSUMPTION

SANOVA TEST #1

Case #	β_1	β_2	PAC	ASN
1	0.00	3.00	0.9006	7.463
2	4.00	12.00	0.9047	7.655
3	4.00	14.00	0.9014	7.695
4	4.00	16.00	0.9023	7.692
5	4.00	18.00	0.9042	7.703
6	4.00	20.00	0.9068	7.716
7	6.25	14.00	0.9060	7.742
8	6.25	16.00	0.9025	7.712
9	6.25	18.00	0.9066	7.740
10	6.25	20.00	0.9073	7.730
11	9.00	24.00	0.9048	7.783
12	25.00	75.00	0.9106	7.938
13	1.00	5.00	0.9018	7.553
14	0.00	4.00	0.9002	7.517
15	0.00	8.00	0.8991	7.626
16	0.25	4.00	0.9026	7.497
17	0.96	4.76	0.8987	7.526
18	2.05	6.85	0.9030	7.600
19	2.25	8.00	0.9000	7.622
20	3.27	9.62	0.9025	7.653

TABLE 14

DEPARTURES FROM NORMALITY ASSUMPTION

SANOVA TEST #3

Case #	β_1	β_2	PAC	ASN
1	0.00	3.00	0.9835	16.47
2	0.00	4.00	0.9828	16.53
3	0.00	6.00	0.9840	16.56
4	0.00	8.00	0.9848	16.63
5	0.25	4.00	0.9840	16.52
6	0.25	6.00	0.9830	16.61
7	0.25	8.00	0.9837	16.61
8	1.00	6.00	0.9840	16.57

As seen from these tables the normality assumption had very little effect on the power of the test at the null hypothesis. As in the fixed sample test, the effect of kurtosis appears to be larger than that of skewness.

4.3 CONCLUSION

This chapter of the thesis has consisted of an investigation of the robustness of sequential analysis of variance tests, specifically, how the power of such a test is affected by deviations from the assumptions. The assumptions considered were homogeneity of variance and normality.

The studies conducted have shown that the SANOVA test is remarkably robust. This might be due in part to the experimental procedure of a SANOVA test, where at each stage of the experiment an equal number of observations from each of the k groups has been taken. And, as studies of the fixed sample ANOVA test have shown, deviations have minimal effects when all group sample sizes are equal.

These findings chould hopefully increase the number of problems for which SANOVA may be useful.

CHAPTER 5

CONCLUSION

This thesis has developed procedures for obtaining the properties of a sequential analysis of variance test (SANOVA). The major properties of interest are the operating characteristic (OC) and average sample number (ASN) curves. These two curves are extremely important since they allow one to design a SANOVA test prior to experimentation.

For a test comparing only two means, this thesis has developed an exact procedure for obtaining the OC and ASN curves. Previous methods have all been approximate, the most common of them being Monte Carlo simulation. My exact procedure uses Aroian's direct method of sequential analysis and has been developed in sufficient detail to serve as a valuable tool for designing experiments.

For SANOVA tests involving more than two means, a new procedure has been developed in which the sequential test probabilities of acceptance, rejection, and continuation are approximated by multivariate normal (MVN) probabilities. This is accomplished by transforming the original SANOVA test statistic and regions so as to (approximately) normalize the test statistic. During

the course of my research in this area the moment generating function of a new type of noncentral bivariate χ^2 distribution was derived. Further research is required to obtain the multivariate extension as well as an expression for the density function. In addition, the ratios of the modified χ^2 variates define a new type of noncentral F distribution.

The MVN approximation gave values in close conformance to those generated by direct Monte Carlo simulation for a variety of parameter sets, and this procedure may eventually supplant Monte Carlo simulations as various approximations for multivariate normal probabilities are further developed. During the investigation of this topic of the thesis one such approximation was developed. Although this was beyond the scope of this thesis, preliminary research has shown that the procedure may be useful.

An experimenter is often concerned whether the assumptions are satisfied for a SANOVA test. Much research has been conducted as to the robustness of fixed sample tests, but relatively few papers have appeared on the robustness of sequential tests. Studies in this thesis revealed that a SANOVA test is robust to deviations in both the normality and homogeneity of variance assumptions. This finding should broaden the experimental situations to which SANOVA is applicable.

REFERENCES

1. Abdel-Aty, S.H. (1954). Approximate Formulae for the Percentage Points and the Probability Integral of the Noncentral χ^2 Distribution, Biometrika 41, 538-40.

2. Abramowitz, M. and Stegun, I.A. (1964). Handbook of Mathematical Functions, National Bureau of Standards.

3. Aroian, L.A. (1941). Continued Fractions for the Incomplete Beta Functions, AMS 12, 218-23. (Correction, AMS 39, 1265).

4. Aroian, L.A. (1941). A study of R.A. Fisher's Z-Distribution and the Related F Distribution, AMS 12, 429-48.

5. Aroian, L.A. (1942). The Relationship of Fisher's Z-Distribution to Student's t Distribution (Abstract), AMS 13, 451-2.

6. Aroian, L.A. (1943). A Certain Type of Integral, American Mathematical Monthly 50, 382-3.

7. Aroian, L.A. (1940). On the Levels of Significance of the Incomplete Beta Function and the F Distributions, Biometrika 37, 219-23.

8. Aroian, L.A. (1976a). Moments of the Decisive Sample Number and of the Average Time to Termination of Sequential Tests, Communications in Statistics 1, 1041-45.

9. Aroian, L.A. (1976b). Applications of the Direct Method in Sequential Analysis, Technometrics 18, 301-6.

10. Aroian, Gorge, Goss, and Robison (1975). Exact Sequential Tests for the Variance, AES Monograph 7501, Union College, Schenectady, New York.

11. Aroian, L.A. and Robison, E.E. (1969). Direct Methods for Exact Truncated Sequential Tests of the Mean of a Normal Distribution, Technometrics 11, 661-75.

12. Bahadur, R.R. (1954). Sufficiency and Statistical Decision Functions, AMS 25, 423-62.

13. Bhate, D.H. (1955). Sequential Analysis with Special Reference to Distributions of Sample Size. Unpublished Ph.D. Thesis, University of London.

14. Bhate, D.H. (1959). Approximation to the Distribution of Sample Size for Sequential Tests. I. Tests of Simple Hypotheses, Biometrika 46, 130-8.

15. Bhattacharjee, G.P. and Nagendra, Y. (1964). Effect of Non-normality on a Sequential Test for Mean, Biometrika 51, 281-7.

16. Box, G.E.P. (1953). Non-normality and Tests on Variances, Biometrika 40, 318-330.

17. Box, G.E.P. (1954). Some Theorems on Quadratic Forms Applied in the Study of the Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification, AMS 25, 290-302.

18. Box, G.E.P. and Andersen, S.L. (1955). Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption, JRSS Series B 17, 1-34.

19. Box, G.E.P. and Muller, M.E. (1958). A Note on the Generation of Normal Deviates, AMS 29, 610-11.

20. Brown, M.B. and Forsythe, A.B. (1974). The Small Sample Behavior of Some Statistics which Test the Equality of Several Means, Technometrics 16, 129-32.

21. Carnahan, et al. (1969). Applied Numerical Methods, Wiley, New York.

22. Chambers, J.M. (1967). On Methods of Asymptotic Approximation for Multivariate Distributions, Biometrika 54, 367-83.

23. Cox, D.R. (1952). Sequential Tests for Composite Hypotheses. Proc. Camb. Phil. Soc. 48, 290-9.

24. Cramier, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J.

25. David, F.N. and Johnson, N.L. (1951). A Method of Investigating the Effect of Non-Normality and Heterogeneity of Variance on Tests of the General Linear Hypothesis, AMS 22, 382-92.

26. Davis, P.J. and Rabinowitz, P. (1967). Numerical Integration, Blaisdell, Waltham, Massachusetts.

27. Eden, T. and Yates, F. (1933). On the Validity of Fisher's Z-Test when Applied to an Actual Sample of Non-normal Data, J. Agricultural Science 23, 6-16.

28. Eisenhardt, C. (1947). The Assumptions Underlying the Analysis of Variance, Biometrics 3, 1-21.

29. Elderton, W.P. and Johnson, N.L. (1969). Systems of Frequency Curves, London, Cambridge University Press.

30. Ewens, W.J. (1961). Departure from Assumption in Sequential Analysis, Biometrika 48, 206-11.

31. Fisher, R.A. (1918). The Correlation Between Relatives on the Supposition of Mendelian Inheritance, Trans. Royal Society of Edinburgh 52, 399-433.

32. Fisher, R.A. (1925). Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh.

33. Fisher, R.A. (1935). The Design of Experiments, Oliver Oliver and Boyd, Edinburgh.

34. Fix, E. (1949). Tables of Noncentral χ^2. University of California Publications in Statistics, Vol. 1, 15-19.

35. Fox, M. (1956). Charts of the Power of the F-Test, AMS 27, 484-97.

36. Gayen, A.K. (1950). The Distribution of the Variance Ratio in Random Samples of Any Size Drawn from Non-normal Universes, Biometrika 37, 236-55.

37. Geary, R.C. (1930). The Frequency Distribution of the Quotient of Two Normal Variates, JRSS 93, 442-6.

38. Geary, R.C. (1936). The Distribution of Student's Ratio for Non-normal Samples, JRSS Series B 3, 178-84.

39. Ghosh, B.K. (1970). Sequential Tests of Statistical Hypotheses, Addison-Wesley, Reading, Massachusetts.

40. Ghosh, B.K. and Eest, J. (1967). Tables for Sequential Analysis of Variance, U.S. Army Report, USA TE Com 7-3-9906-01, Fort Lee, Virginia.

41. Gradshteyn, I.S. and Ryshik, I.M. (1965). Tables of Integrals, Series, and Products, Academic Press, New York.

42. Graybill, F.A. (1961). An Introduction to Linear Statistical Models, Vol. 1, McGraw-Hill, New York.

43. Graybill, F.A. (1969). Introduction to Matrices with Applications in Statistics, Wadsworth Publishing Company, Inc., Belmont, California.

44. Gronow, D.G.C. (1951). Test for the Significance of the Difference Between Means in Two Normal Populations Having Unequal Variances, Biometrika 38, 252-56.

45. Gulberg, S. (1920). Application des Polynômes d'Hermite à un Problème de Statistique, Proceedings of the International Congress of Mathematicians, Strasbourg, 552-60.

46. Gurland, J. (1948). Inversion Formulae for the Distribution of Ratios, AMS 19, 228-37.

47. Gurland, J. (1953). Distribution of Quadratic Forms and Ratios of Quadratic Forms, AMS 26, 122-27.

48. Hahn, G.J. and Shapiro, S. (1969). Statistical Models in Engineering, Wiley, New York.

49. Hall, W.J., Wijsman, R.A., and Ghosh, J.K. (1965). The Relationship between Sufficiency and Invariance with Applications in Sequential Analysis, AMS 36, 574-614.

50. Helmert, F.R. (1876). Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und Über einige damit in Zusammenhänge stehende Fragen, Zeitschrift für Angewandte Mathematik und Physik 21, 192-218.

51. Hoel, P.G. (1955). On a Sequential Test for the General Linear Hypothesis, AMS 26, 136-48.

52. Horsnell, G. (1953). The Effect of Unequal Group Variances on the F-test for Homogeneity of Group Means, Biometrika 40, 128-36.

53. Hsu, P.L. (1941). Analysis of Variance from the Power Function Standpoint, Biometrika 32, 62-9.

54. Johnson, N.L. (1948). Alternative Systems in the Analysis of Variance, Biometrika 35, 80-87.

55. Johnson, N.L. (1949). Systems of Frequency Curves Generated by Methods of Truncation, Biometrika 36, 149-76.

56. Johnson, N.L. (1953). Some Notes on the Application of Sequential Methods in the Analysis of Variance, AMS 25, 357-66.

57. Johnson, N.L. and Kotz, S. (1969). Distributions in Statistics: Discrete Distributions, Wiley, New York.

58. Johnson, N.L. and Kotz, S. (1970). Distributions in Statistics: Continuous Univariate Distributions-1, Wiley, New York.

59. Johnson, N.L. and Kotz, S. (1970). Distributions in Statistics: Continuous Univariate Distributions-2, Wiley, New York.

60. Johnson, N.L. and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions, Wiley, New York.

61. Kempthorne, O. (1952). Design and Analysis of Experiments, Hafner, New York.

62. Kendall, M.G. and Stuart, A. (1963). The Advanced Theory of Statistics, Vol. 1, Hafner, New York.

63. Kendall, M.G. and Stuart, A. (1967). The Advanced Theory of Statistics, Vol. 2, Hafner, New York.

64. Kibble, W.F. (1941). A Two-Variate Gamma Type Distribution, Sankhya 5, 137-50.

65. Kohr, R.L. and Games, P.A. (1974). Robustness of the Analysis of Variance, The Welch Procedure and a Box Procedure to Heterogeneous Variances, The Journal of Experimental Education 43, 61-68.

66. Kolodziejczyk, S. (1935). On an Important Class of Statistical Hypotheses, Biometrika 27, 161-90.

67. Lehman, E.L. (1959). Testing Statistical Hypotheses, Wiley, New York.

68. Lehmer, E. (1944). Inverse Tables of Probabilities of Errors of the Second Kind, AMS 15, 388-98.

69. Meixner, J. (1934). Orthogonale Polynomsysteme mit emer besonderen Gestalt des erzeugenden Funktion, Journal of the London Mathematical Society 9, 6-13.

70. Mudholkar, G.S., et al. (1976). Some Approximations for the Noncentral F Distribution, Technometrics 18, 351-8.

71. Naylor, T.H., et al. (1966). Computer Simulation Techniques, Wiley, New York.

72. Patnaik, P.B. (1949). The Noncentral χ^2 and F Distributions and Their Applications, Biometrika 36, 202-32.

73. Paulson, E. (1942). An Approximate Normalization of the Analysis of Variance Distribution, AMS 13, 233-35.

74. Pearson, E.S. (1929). The Distribution of Frequency Constants in Small Samples from Non-normal Symmetrical and Skew Populations, Biometrika 21, 259-86.

75. Pearson, E.S. (1931). The Analysis of Variance in Cases of Non-normal Variation, Biometrika 23, 114-33.

76. Pearson, E.S. and Hartley, H.O. (1951). Charts of the Power Function of the Analysis of Variance Tests, Derived from the Non-central F-Distribution, Biometrika 38, 112-30.

77. Pierce, D.A. and Dykstra, R.L. (1969). Independence and the Normal Distribution, American Statistician 23, No. 4, 39.

78. Press, S.J. (1966). Linear Combination of Noncentral Chi-Square Variates, AMS 37, 480-87.

79. Ray, W.D. (1956). Sequential Analysis Applied to Certain Experimental Designs in the Analysis of Variance, Biometrika 43, 388-403.

80. Robbins, H.E. and Pitman, E.J.G. (1949). Applications of the Methods of Mixtures to Quadratic Forms in Normal Variates, AMS 20, 552-60.

81. Sankaran, M. (1963). Approximations to the Noncentral Chi-Square Distribution, Biometrika 50, 199-204.

82. Scheffé, H. (1942). On the Ratio of Variances of Two Normal Populations, AMS 13, 371-88.

83. Scheffé, H. (1959). The Analysis of Variance, Wiley, New York.

84. Schmee, J. (1974). Exact Solution for the Sequential t-Test and Method for Sequential Estimation, Ph.D. dissertation, Union College, Schenectady, New York.

85. Siegmund, D. (1976). Importance Sampling in the Monte Carlo Study of Sequential Tests, Annals of Statistics 4, 673-84.

86. Slater, L.J. (1960). Confluent Hypergeometric Functions, Cambridge University Press.

87. Srivastava, A.B.L. (1958). Effect of Non-normality on the Power Function of the t-Test, Biometrika 45, 421-9.

88. Subrahmaniam, K. (1966). Some Contributions to the Theory of Non-normality-I (Univariate Case), Sankhya, Series A 28, 389-406.

89. Subrahmaniam, K. (1968). Some Contributions to the Theory of Non-normality-II, Sankhya, Series A 30, 411-32.

90. Tang, P.C. (1938). The Power Function of the Anslysis of Variance Tests with Tables and Illustrations of their Use, Statistical Research Memoirs 2, 126-49.

91. Tiku, M.L. (1964). Approximating the General Non-normal Variance-Ratio Sampling Distribution, Biometrika 51, 83-95.

92. Tiku, M.L. (1965). Laguerre Series Forms of Noncentral χ^2 and F Distributions, Biometrika 52, 415-27.

93. Tiku, M.L. (1966). A Note on Approximating the Noncentral F Distribution, Biometrika 53, 606-10.

94. Wald, A. (1942a). On the Power Function of the Analysis of Variance Test, AMS 13, 96-100.

95. Wald, A. (1942b). On the Principles of Statistical Inference, Notre Dame Math. Lecture, No. 1, Edwards Brothers, Ann Arbor, Michigan.

96. Wald, A. (1947). Sequential Analysis, Wiley, New York.

97. Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test, AMS 19, 327-39.

98. Wallace, P.L. (1959). Bounds on Normal Approximations to Student's and the Chi-Square Distributions, AMS 30, 1121-30.

99. Wetherill, G.B. (1966). Sequential Methods in Statistics, Wiley, New York.

100. Wilson, E.B. and Hilferty, M.M. (1931). The Distribution of Chi-Square, Proceedings of the National Academy of Science, Washington, 17, 684-88.

APPENDIX D

DERIVATIONS NECESSARY FOR THE MULTIVARIATE NORMAL APPROXIMATION

D.1 DERIVATION OF THE JOINT DISTRIBUTION OF D_n AND D_m

As discussed in Section (3.2) of the thesis, the SANOVA test uses the following V_n statistic:

$$V_n = \frac{T_n}{D_n} = \frac{n \sum_{i=1}^{K} (\bar{X}_{i(n)} - \bar{X}_{(n)})^2 / \sigma^2}{\sum_{i=1}^{K} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{i(n)})^2 / \sigma^2}$$

The statistics D_n and D_m of any two stages n and m $(m > n)$ are not independent. In fact

$$D_n = \sum_{i=1}^{K} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{i(n)})^2 / \sigma^2$$

$$D_m = \sum_{i=1}^{K} \sum_{j=1}^{n} (X_{ij} - \bar{X}_{i(m)})^2 / \sigma^2$$

$$= D_n + \left\{ n \sum_{i=1}^{K} (\bar{X}_{i(m)} - \bar{X}_{i(n)})^2 + \sum_{i=1}^{K} \sum_{j=n+1}^{m} (X_{ij} - \bar{X}_{i(m)})^2 \right\} / \sigma^2$$

$$= D_n + W$$

It can easily be shown that each of these quantities is marginally distributed as:

$$D_n \sim \chi^2(\nu_{2n})$$

$$W \sim \chi^2(\nu_2)$$

$$D_m \sim \chi^2(\nu_{2m})$$

where

$$\nu_{2n} = K(n-1)$$

$$\nu_2 = K(m-n)$$

$$\nu_{2m} = K(m-1)$$

Since W and D_n are independent, their joint distribution is simply the product of two χ^2 densities:

$$f(W,D_n) = \frac{W^{\frac{1}{2}(\nu_2-2)}\ D_n^{\frac{1}{2}(\nu_{2n}-2)}\ e^{-\frac{1}{2}(W+Dn)}}{\Gamma(\nu_{2n}/2)\ \Gamma(\nu_2/2)\ 2^{(\nu_2+\frac{1}{2}\nu_{2n})}}$$

Since

$$W = D_m - D_n ,$$

the joint density of D_n and D_m is given by:

$$f(D_n,D_m) = \frac{(D_m-D_n)^{\frac{1}{2}(\nu_2-2)}\ D_n^{\frac{1}{2}(\nu_{2n}-2)}\ e^{-\frac{1}{2}D_m}}{\Gamma(\nu_{2n}/2)\ \Gamma(\nu_2/2)\ 2^{(\nu_2+\ \nu_{2n})}} ,$$

which upon substituting the values for ν_2 and ν_{2n} becomes:

$$f(D_n, D_m) = \frac{(D_m - D_n)^{\frac{1}{2}(K(m-n)-2)} \; D_n^{\frac{1}{2}(K(n-1)-2)} \; e^{-\frac{1}{2}D_m}}{(K(n-1)/2) \quad (K(m-n)/2) \quad 2^{\frac{1}{2}K(m-1)}}$$

$$\text{for } D_m \geq D_n \tag{D.1.1}$$

The joint raw moments are obtained by performing the following integration of the density of equation (D.1.1):

$$E\left[D_n^{\,r} \; D_m^{\,s}\right] = \int_0^{\infty} \int_0^{D_m} D_n^{\,r} \, D_m^{\,s} \, f(D_n, D_m \; dD_n \; dD_m$$

The result of this integration yields the following closed form expression:

$$E\left[D_n^{\,r} \; D_m^{\,s}\right] = \frac{\Gamma(r+\frac{1}{2}K(n-1)) \;\; \Gamma(r+s+\frac{1}{2}K(m-1))(2^{r+s})}{\Gamma(r+\frac{1}{2}K(m-1)) \;\; \Gamma(\frac{1}{2}K(n-1))} \tag{D.1.2}$$

D.2 DERIVATION OF THE JOINT MOMENT GENERATING FUNCTION OF D_n AND D_m

The joint moment generating function of D_n and D_m may be obtained by performing the following integration:

$$M_{D_n,D_m}(t_1,t_2) = \int_0^{\infty}\int_0^{D_m} f(D_n,D_m)\, e^{t_1 D_n}\, e^{t_2 D_m}\, dD_n\, dD_m$$

Upon substitution this becomes the following:

$$M_{D_n,D_m}(t_1,t_2) = \left[\Gamma(\nu_2/2)\ \Gamma(\nu_{2n}/2)\ 2^{\frac{1}{2}(\nu_2+\nu_{2n})}\right]^{-1}$$

$$\int_0^{\infty} e^{-\frac{1}{2}D_m}\, e^{t_2 D_m} \int_0^{D_m} D_n^{\frac{1}{2}(\nu_2-2)}\, (D_m-D_n)^{\frac{1}{2}(\nu_{2n}-2)}\, e^{t_1 D_n}\, dD_n\, dD_m$$

The first integral is given in Gradshteyn and Ryzhik (1965) (p. 318, integral #3.383) as:

$$M_{D_n',D_m}(t_1,t_2) = \left[\Gamma(\nu_2/2)\ \Gamma(\nu_{2n}/2)\ 2^{\frac{1}{2}(\nu_2+\nu_2 n)}\right]^{-1}$$

$$\int_0^\infty e^{-\frac{1}{2}D_m}\ e^{t_2 D_m}\ B(\tfrac{1}{2}\nu_2,\ \tfrac{1}{2}\nu_{2n})\ D_m^{\frac{1}{2}(\nu_2+\nu_{2n}-2)}$$

$${}_1F_1(\tfrac{1}{2}\nu_2, \tfrac{1}{2}(\nu_2+\nu_{2n}),\ t_1 D_m\ \ dD_m$$

where

$$B(X,Y) = \Gamma(X)\Gamma(Y)\ /\ \Gamma(X+Y)$$

and

$${}_1F_1(X,Y,Z) = \sum_{i=0}^{\infty} \frac{\Gamma(X+i)\ \Gamma(Y)}{\Gamma(Y+i)\ \Gamma(X)}\ \frac{Z^i}{i!}$$

The remaining integral yields (Gradshteyn and Ryzhik (1965), p. 860, #7.6214):

$$M_{D_n,D_m}(t_1,t_2) = \left[\Gamma(\nu_2/2)\ \Gamma(\nu_{2n}/2)\ 2^{\frac{1}{2}(\nu_2+\nu_{2n})}\right]^{-1}$$

$$B(\nu_2/2,\ \nu_{2n}/2)\ \Gamma(\tfrac{1}{2}(\nu_2+\nu_{2n}))\ (\tfrac{1}{2}-t_2)^{-\frac{1}{2}(\nu_2+\nu_{2n})}$$

$${}_2F_1\left(\nu_2/2,\ \tfrac{1}{2}(\nu_2+\nu_{2n}),\ \tfrac{1}{2}(\nu_2+\nu_{2n}),\ t_1(\tfrac{1}{2}-t_2)^{-1}\right)$$

where ${}_2F_1(a,b,c,Z)$ is the Gauss Hypergeometric function (Slater (1965)).

The above expression can be simplified to the following form:

$$M_{D_n,D_m}(t_1,t_2) = (1-2t_2)^{-\frac{1}{2}(\nu_2+\nu_{2n})}$$

$${}_2F_1(\nu_2/2,\ \tfrac{1}{2}(\nu_2+\nu_{2n}),\ \tfrac{1}{2}(\nu_2+\nu_{2n}),\ t_1(\nu-t_2)^{-1}$$

This may be further simplified by noting the following identity (Abramowitz and Stegun (1964)):

$$ {}_2F_1(x,y,y,Z) \;=\; (1-Z)^{-X} $$

Performing this substitution and simplifying yields the following expression

$$ M_{D_n,D_m}(t_1,t_2) \;=\; (1-2t_2)^{-\frac{1}{2}\nu_{2n}}\,(1-2t_1-2t_2)^{-\frac{1}{2}\nu_2} $$

$$ = (1-2t_2)^{-\frac{1}{2}K(n-1)}\,(1-2t_1-2t_2)^{-\frac{1}{2}K(m-n)} \qquad \text{(D.2.1)} $$

D.3 DERIVATION OF THE MIXED CENTRAL MOMENTS OF D_n AND D_m

The previous sections of this appendix have derived expressions for: the joint density of D_n and D_m; the mixed raw moments of D_n and D_m; and the joint moment generating function of D_n and D_m. All three of these expressions are useful for obtaining the mixed central moment of D_n and D_m.

The mixed central moments are defined as:

$$E\left[(D_n-\mu_{D_n})^r\ (D_m-\mu_{D_m})^s\right];$$

and are needed for the Taylor series approximation to cov (Z_n, Z_m) of Section 3.2.

The covariance is obtained as

$$\text{cov}\ (D_n D_m) = E\left[D_n D_m\right] - E\left[D_n\right] E\left[D_m\right]$$

which upon substituting the results of equation (D.1.2) yields

$$\begin{aligned}\text{cov}\ (D_n D_m) &= 2K(n-1) + K^2(m-1)(n-1) - K^2(m-1)(n-1) \\ &= 2K(n-1)\end{aligned} \tag{D.3.1}$$

Similar expressions were obtained for the higher central moments and are summarized in Table 15.

TABLE 15

MIXED CENTRAL MOMENTS OF D_n AND D_m

$$E\left[(D_n-\mu_{D_n})^r\ (D_m-\mu_{D_m})^s\right]$$

r	s	Mixed Moment
0	2	2K(m-1)
0	3	8K(m-1)
1	1	2K(n-1)
1	2	8K(n-1)
2	0	2K(n-1)
2	1	8K(n-1)
3	0	8K(n-1)

D.4. DERIVATION OF THE JOINT MOMENT GENERATING FUNCTION OF T_n AND T_m

The statistic T_n is given by

$$T_n = n \sum_{i=1}^{K} (\bar{X}_{i(n)} - \bar{\bar{X}}_{(n)})^2/\sigma^2$$

where

$$\bar{X}_{i(n)} = \sum_{j=1}^{n} X_{ij}/n$$

and

$$\bar{\bar{X}}_{(n)} = \sum_{i=1}^{K} X_{i(n)}/K$$

The joint moment generating function of the statistics T_n and T_m is best derived by using the following set of transformations:

$$U_1 = (\bar{X}_{1(n)} + \bar{X}_{2(n)})/\sqrt{2}\,\sigma$$

$$U_2 = (\bar{X}_{1(n)} + \bar{X}_{2(n)} - 2\bar{X}_{3(n)})/\sqrt{6}\,\sigma$$

$$\vdots$$

$$U_{K-1} = (\bar{X}_{1(n)} + \bar{X}_{2(n)} + \cdots + X_{K-1(n)} - (K-1)\,\bar{X}_{K(n)})/\sqrt{K(K-1)}\,\sigma$$

and

$$V_i = (\bar{X}_{1(m)} - \bar{X}_{2(m)})/\sqrt{2}\,\sigma$$

$$V_2 = (\bar{X}_{1(m)} + \bar{X}_{2(m)} - 2\bar{X}_{3(m)}/\sqrt{6}\,\sigma$$

$$\vdots$$

$$V_{K-1} = (\bar{X}_{1(m)} + \bar{X}_{2(m)} + \cdots + \bar{X}_{K-1(m)} - (K-1)\bar{X}_{K(m)})\,/\sqrt{K(K-1)}\,\sigma \tag{D.4.1}$$

These transformations are those used by Helmert (1876); and are defined so that:

$$\sum_{i=1}^{K-1} u_i^2 = \sum_{i=1}^{K} (\bar{X}_{i(n)} - \bar{\bar{X}}_{(n)})^2/\sigma^2$$

$$\sum_{i=1}^{K-1} v_i^2 = \sum_{i=1}^{K} (\bar{X}_{i(m)} - \bar{\bar{X}}_{(m)})^2/\sigma^2 \tag{D.4.2}$$

In addition it can easily be shown that:

$\mathrm{var}(U_i) = 1/n$

$\mathrm{var}(V_i) = 1/m$

$\mathrm{cov}(U_i, U_j) = 0$

$\mathrm{cov}(V_i, V_j = 0$

Also of interest is $\mathrm{cov}(U_i, V_j)$. This can be obtained by realizing that:

$$V_j = (\bar{X}_{1(m)} + \bar{X}_{2(m)} + \cdots \bar{X}_{j-1(m)} - (j-1)\bar{X}_{j(m)}) / \sqrt{j(j-1)}$$

$$= \frac{1}{m}\Bigg\{ n\bar{X}_{1(n)} + n\bar{X}_{2(n)} + \cdots + n\bar{X}_{j-1(n)} - n(j-1)\bar{X}_{j(n)}$$

$$+ \sum_{i-1}^{j-Z} \sum_{\ell=n+1}^{m} X_{i\ell} - (j-1) \sum_{\ell=n+1}^{m} X_{j\ell} \Bigg\} / \sqrt{j(j-1)}$$

$$= \frac{1}{m} \left\{ nU_j + E_j \right\} .$$

then

$$\mathrm{cov}(U_i, V_j) = E\left[U_i V_j\right] - E\left[U_i\right] E\left[V_j\right]$$

$$= E\left[U_i \left(\frac{1}{m} \left\{ nU_j + E_j \right\}\right)\right] - E\left[U_i\right] E\left[\frac{1}{m}(nU_j + E_j)\right]$$

$$= \frac{n}{m} E\left[U_i U_j\right] + \frac{1}{m} E\left[U_i\right] E\left[E_j\right] - \frac{n}{m} E\left[U_i\right] E\left[U_j\right]$$

$$- \frac{1}{m} E\left[U_i\right] E\left[E_j\right]$$

$$= \frac{n}{m}\left\{ E\left[U_i U_j\right] - E\left[U_i\right] E\left[U_j\right] \right\} . \tag{D.4.2}$$

when $i \neq j$

$$\mathrm{cov}(U_i, V_j) = \frac{n}{m}\left\{E[U_i]\,E[U_j] - E[U_i]\,E[U_j]\right\} = 0$$

since U_i and U_j are independent.

However, when $i = j$

$$\mathrm{cov}(U_i, V_j) = \frac{n}{m}\left\{E[U_i^2] - \left(E[U_i]\right)^2\right\}$$

$$= \frac{n}{m}\,\mathrm{VAR}(U_i) = \left(\frac{n}{m}\right)\left(\frac{1}{n}\right) = \frac{1}{m}$$

Therefore, whenever $\overline{X}_{i(n)}$, $\overline{X}_{i(m)}$ are normally distributed; the vector X, where

$$X' = \left[U_1 U_2 \cdots U_{K-1} V_1 V_2 \cdots V_{K-1}\right] \tag{D.4.3}$$

will be multivariate normally distributed with

$$E\left[X\right] = W \tag{D.4.4}$$

where

$$W_1 = (\mu_1 - \mu_2) / \sqrt{2}\,\sigma$$

$$W_2 = (\mu_1 + \mu_2 - 2\mu_3) / \sqrt{6}\,\sigma$$

$$W_{K-1} = (\mu_1 + \mu_2 + \cdots + \mu_{K-1} - (K-1)\mu_K) / \sqrt{K(K-1)}\,\sigma$$

$$W_K = (\mu_1 - \mu_2) / \sqrt{2}\,\sigma$$

$$\vdots$$

$$W_{2K-1} = (\mu_1 + \mu_2 + \cdots + \mu_{K-1} - (K-1)\mu_K) / \sqrt{K(K-1)}\,\sigma$$

and variance covariance matrix Σ; where Σ can be expressed as the following partitioned matrix:

$$\Sigma = \left[\begin{array}{c|c} \frac{1}{n} I & \frac{1}{m} I \\ \hline \frac{1}{m} I & \frac{1}{m} I \end{array}\right] \tag{D.4.5}$$

I being a K-1 x K-1 identity matrix.

The density of X is given by

$$f(X) = (2\pi)^{-(K-1)} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(X-W)' \Sigma^{-1} (X-W)\right\} .$$

The joint moment generating function of T_n and T_m may then be obtained by performing the following multi-dimensional integration:

$$M_{T_n,T_m}(t_1,t_2) = (2\pi)^{-(K-1)} |\Sigma|^{-\frac{1}{2}}$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{t_1 X'AX\right\} \exp\left\{t_2 X'BX\right\} \exp\left\{-\tfrac{1}{2}(X-W)'\Sigma^{-1}(X-W)\right\} dX_1 \cdots dX_{2K-2} \tag{D.4.6}$$

where A and B are matrices which may be expressed in the following block partitioned forms:

$$A:\quad 2K-2 \times 2K-2 \qquad \left[\begin{array}{c|c} nI & \phi \\ \hline \phi & \phi \end{array}\right]$$

$$B:\quad 2K-2 \times 2K-2 \qquad \left[\begin{array}{c|c} \phi & \phi \\ \hline \phi & mI \end{array}\right]$$

(D.4.7)

where I is again the $K-1 \times K-1$ identity matrix and ϕ is the $K-1 \times K-1$ null matrix.

The integration of (D.4.6) is performed in Graybill (1969) (p. 252) and yields

$$M_{T_n,T_m}(t_1,t_2) = \left|I-2t_1\sum A-2t_2\sum B\right|^{-\frac{1}{2}}$$

$$\exp\left\{(\tfrac{1}{2})(\textstyle\sum^{-1}W)'(\sum^{-1}-2t\ A-2t\ B)^{-1}(\sum^{-1}W)(-1)(W')\sum^{-1}W\right\}$$

(D.4.8)

The determinant portion may be simplified to the following block partitioned form:

$$\left| I-2t_1\sum A-2t_2\sum B \right| = \left\|\left[\begin{array}{c|c} (1-2t_1)\ I & -2t_2\ I \\ \hline -2t_1(\frac{n}{m})\ I & (1-2t_2)\ I \end{array}\right]\right\|,$$

which yields Graybill (1969) (p. 165)

$$= (1-2t_1)^{k-1} \left| (1-2t_2)I - \frac{4t_1t_2n}{(1-2t_1)m} \right|$$

$$= \left\{ (1-2t_1)(1-2t_2) - 4nt_1t_2m^{-1} \right\}^{-\frac{1}{2}(K-1)} \tag{D.4.9}$$

This expression is similar to that obtained by Kibble (1945) in his derivation of a moment generating function for a type of bivariate gamma distribution.

The exponent may be rearranged as follows:

$$\left(\Sigma^{-1}W\right)'\left(\Sigma^{-1}-2t_1A-2t_2B\right)^{-1}\left(\Sigma^{-1}W\right)\left(-W\right)\Sigma^{-1}W$$

$$= W'\Sigma^{-1}\left[\left(I-2t_1\Sigma A-2t_2\Sigma B\right)^{-1}-I\right]W$$

The inner product can be simplified to the following block partitioned matrix

$$\Sigma^{-1}\left[\left(I-2t_1\Sigma A-2t_2\Sigma B\right)^{-1}-I\right] = \alpha \left[\begin{array}{c|c} r_{11}I & r_{12}I \\ \hline r_{21}I & r_{22}I \end{array}\right] = R$$

where

$$r_{11} = m^2n(1-2t_2)-m^2n(1-2t_1)(1-2t_2)+4mn^2t_1t_2-2mn^2t_1$$

$$r_{12} = 2m^2nt_2-m^2n(1-2t_1)+m^2n(1-2t_1)(1-2t_2)-4t_1t_2mn^2$$

$$r_{21} = 2m^2nt_1-m^2n(1-2t_2)+m^2n(1-2t_1)(1-2t_2)-4t_1t_2mn^2$$

$$r_{22} = m^3(1-2t_1)-m^3(1-2t_1)(1-2t_2)+4t_1t_2m^2n-2m^2nt_2$$

$$\alpha = \left\{\left[m(1-2t_1)(1-2t_2)-4t_1t_2n\right](m-n)\right\}^{-1}$$

Now, the vector W is symmetrical in that the first $K-1$ elements are identical to the second $K-1$ elements. Denote this symmetry by

$$W' = \left[w'|w'\right]$$

The exponent is then given by

$$W'RW = \alpha\left[w'|w'\right]\left[\begin{array}{c|c} r_{11}I & r_{12}I \\ \hline r_{21}I & r_{22}I \end{array}\right]\left[\begin{array}{c} W \\ \hline W \end{array}\right]$$

$$= \alpha\, W'W(r_{11}+r_{12}+r_{21}+r_{22})$$

The quantity $W'W$ is simply λ, so the above becomes

$$= \frac{\lambda\left\{2nt_1+2mt_2-4t_1t_2(m-n)\right\}}{\left[(1-2t_1)(1-2t_2)-4t_1t_2nm^{-1}\right]}\ .$$

Thus, the moment generating function is given by the following expression:

$$M_{T_n, T_m}(t_1, t_2) = \left\{(1-2t_1)(1-2t_2) - 4t_1 t_2 n m^{-1}\right\}^{-\frac{1}{2}(K-1)}$$

$$\exp\left\{\frac{\lambda\left[2nt_1 + 2mt_2 - 4t_1 t_2 (m-n)\right]}{2\left[(1-2t_1)(1-2t_2) - 4t_1 t_2 n m^{-1}\right]}\right\} \qquad \text{(D.4.10)}$$

This is the moment generating function of a type of noncentral bivariate χ^2 distribution.

D.5 THE MIXED CENTRAL MOMENTS OF T_n AND T_m

The mixed central mements of T_n and T_m; i.e.:

$$E\left[(T_n-\mu_{T_n})^r (T_m-\mu_{T_m})^s\right],$$

are best obtained from the joint cumulant generating function of T_n and T_m. The joint cumulant generating function is obtained as the logarithm of the moment generating function given in equation (D.4.10) of the previous section. This yields the following expression:

$$K_{T_n,T_m}(t_1,t_2) = \log\left((M_{T_n,T_m}(t_1,t_2)\right)$$

$$= -\tfrac{1}{2}(K-1)\log\left[(1-2t_1)(1-2t_2)-4t_1t_2nm^{-1}\right]$$

$$+ \lambda\left[nt_1+mt_2-2t_1t_2(m-n)\right]\left[(1-2t_1)(1-2t_2)-4t_1t_2nm^{-1}\right]^{-1} \qquad \text{(D.5.1)}$$

By differentiating or expanding this function one can obtain the joint cumulants. In particular

$$\operatorname{cov}(T_n,T_m) = \left.\frac{\partial^2 K_{T_n,T_m}(t_1,t_2)}{\partial t_1 \partial t_2}\right|_{\substack{t_1=0\\ t_2=0}}$$

$$E\left[\left(T_n-\mu_{T_n}\right)\left(T_m-\mu_{T_m}\right)^2\right] = \left.\frac{\partial^3 K_{T_n,Tm}(t_1,t_2}{\partial t_1 \partial t_2^{\,2}}\right|_{\substack{t_1=0\\ t_2=0}}$$

$$E\left[\left(T_n-\mu_{T_n}\right)^2\left(T_m-\mu T_m\right)\right] = \left.\frac{\partial^3 K_{T_n,T_m}(t_1,t_2)}{\partial t_1^{\,2} \partial t_2}\right|_{\substack{t_1=0\\ t_2=0}}$$

These results are summarized in Table 16.

TABLE 16

MIXED CENTRAL MOMENTS OF T_n AND T_m

$$E\ (T_n-\mu_{T_n})^r (T_m-\mu_{T_m})^s$$

r	s	Moment
0	2	$2(K-1+2m\lambda)$
0	3	$8(K-1+3m\lambda)$
1	1	$2(K-1)nm^{-1}+4n\lambda$
1	2	$24n\lambda+4(K-1)(3nm^{-1}-1)$
2	0	$2(K-1+2n\lambda)$
2	1	$8n(k-1)m^{-1}+24n\lambda-8(m-n)\lambda m^{-1}$
3	0	$8(k-1+3n\lambda)$

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 14 AES-7907	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 Sequential Analysis of Variance: Monte Carlo Simulation, Multivariate Normal Approximation, and Robustness.		5. TYPE OF REPORT & PERIOD COVERED Final. June 1, 1977-Aug. 1, 1979
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) 10 Robert W. Miller		8. CONTRACT OR GRANT NUMBER(s) NR 042-383 15 N00014-77-C-0438
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute of Administration & Management, Union College & University, Schenectady, NY 12308		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 11 1 Aug 79
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of Navy, Arlington, VA		12. REPORT DATE Aug. 1, 1979
		13. NUMBER OF PAGES 162
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 160 p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

9 Final rept. 1 Jun 77-1 Aug 79,

18. SUPPLEMENTARY NOTES

Chapters 1,3,4,5 of doctoral dissertation at Union College and University: Sequential Analysis of Variance.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Direct method, sequential analysis, exact test, analysis of variance, numerical methods, invariant sequential test, approximation methods, Monte Carlo, multivariate normal, robust.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Monte Carlo simulation and a multivariate normal approximation are used to find the OC and ASN of the sequential analysis of variance test. It is shown that the SANOVA test is remarkably robust.

418 725

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)